

JULIANO FABIANO DA MOTA

**UM ESTUDO DE CASO PARA A DETERMINAÇÃO DO PREÇO DE VENDA DE
IMÓVEIS URBANOS VIA REDES NEURAIS ARTIFICIAIS E MÉTODOS
ESTATÍSTICOS MULTIVARIADOS**

CURITIBA

2007

JULIANO FABIANO DA MOTA

**UM ESTUDO DE CASO PARA A DETERMINAÇÃO DO PREÇO DE VENDA DE
IMÓVEIS URBANOS VIA REDES NEURAIS ARTIFICIAIS E MÉTODOS
ESTATÍSTICOS MULTIVARIADOS**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciências pelo Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Área de Concentração Programação Matemática, Setor de Ciências Exatas e de Tecnologia, Universidade Federal do Paraná.

Orientação: Prof^ª. Dr^ª. Maria Teresinha Arns Steiner

CURITIBA

2007

TERMO DE APROVAÇÃO

JULIANO FABIANO DA MOTA

“Um estudo de caso para a determinação do preço de venda de imóveis urbanos via redes neurais artificiais e métodos estatísticos multivariados”

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Curso de Pós-Graduação em Métodos Numéricos em Engenharia – Área de Concentração em Programação Matemática, Setores de Ciências Exatas Tecnologia e de da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientadora:

Prof^a. Maria Teresinha Arns Steiner, D. Eng.
Departamento de Matemática, UFPR.

Prof. Jair Mendes Marques, D. Sc.
Programa de Pós-Graduação em Métodos
Numéricos em Engenharia, UFPR.

Prof. Júlio Cesar Nievola, Ph. D.
Programa de Pós-graduação em Informática
Aplicada, PUC-PR.

Curitiba, 17 de agosto de 2007.

A todo aquele que promove a paz.

AGRADECIMENTOS

A Deus, minha família e à minha namorada, Nádia, pelo apoio e compreensão.

A todos os alunos do curso pelo apoio e amizade e em especial aos colegas Dirceu, Éder, Adriano, Gislaine e Solange pelo companheirismo em todas as disciplinas.

Aos professores do curso, por toda experiência, conhecimento passados e, também, pela amizade.

À minha orientadora, prof^a. Maria Teresinha, que apoiou (com muita paciência) o desenvolvimento do trabalho e ofereceu todas as contribuições necessárias para sua realização.

A todos que direta ou indiretamente contribuíram para a realização deste trabalho.

Não encontre defeitos, encontre soluções! Qualquer um sabe queixar-se.
(Henry Ford)

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE TABELAS	viii
RESUMO	ix
ABSTRACT	x
1 INTRODUÇÃO	1
1.1 PROBLEMA	1
1.2 OBJETIVOS	3
1.2.1 Objetivo Geral	3
1.2.2 Objetivos Específicos	4
1.3 JUSTIFICATIVA	4
1.4 ESTRUTURA DO TRABALHO	5
2 DESCRIÇÃO DO PROBLEMA	6
2.1 A ÁREA DO ESTUDO	6
2.1.2 Limitações da Pesquisa.....	7
2.1.3 Levantamento dos Dados	8
2.1.3.1 As variáveis utilizadas	8
2.1.3.2 Amostra	12
2.2 METODOLOGIA PARA O DESENVOLVIMENTO DA PESQUISA.....	12
3 REVISÃO DE LITERATURA	14
3.1 ENGENHARIA DE AVALIAÇÃO.....	14
3.1.1 Histórico Sobre as Normas Técnicas.....	14
3.1.1.1 Valor	15
3.1.2 Avaliação e Seus Métodos.....	17
3.1.2.1 Método comparativo direto de dados	17
3.1.2.2 Método involutivo	18
3.1.2.3 Método evolutivo.....	18
3.1.2.4 Método da capitalização da renda	18
3.1.3 Diretrizes Para a Aplicação dos Métodos.....	18
3.1.4 Classificação dos Imóveis Urbanos.....	19
3.1.5 O Mercado	20
3.1.5.1 O mercado imobiliário.....	20
3.2 REDES NEURAIIS ARTIFICIAIS	23
3.2.1 Um Breve Histórico.....	23
3.2.2 Características das Redes Neurais Artificiais	24
3.2.2.1 Funções de ativação.....	25
3.2.2.2 Os tipos de topologia das redes neurais.....	27
3.2.2.3 O aprendizado das redes neurais artificiais	29
3.2.3 O Algoritmo <i>Back-propagation</i>	29
3.2.3.1 O algoritmo de treinamento <i>back-propagation</i>	30
3.2.3.2 Os parâmetros utilizados no algoritmo <i>back-propagation</i>	33
3.2.3.3 Variações no algoritmo <i>back-propagation</i>	34
3.2.4 Redes Competitivas	35
3.2.4.1 Redes de Kohonen	36
3.2.4.1.1 Tipos de distância utilizados.....	39
3.3 MÉTODOS ESTATÍSTICOS MULTIVARIADOS	40

3.3.1 Introdução	40
3.3.2 Estatísticas Descritivas Multivariadas	41
3.3.2.1 Padronização de dados multivariada	43
3.3.3 Análise de Componentes Principais	43
3.3.3.1 Componentes principais populacionais	44
3.3.3.2 Componentes principais amostrais	45
3.3.3.3 Critérios para definição do número de componentes principais.....	46
4 METODOLOGIA PARA RESOLUÇÃO DO PROBLEMA	48
4.1 DESCRIÇÃO DO SISTEMA AIRNA	48
5 OBTENÇÃO E ANÁLISE DOS RESULTADOS	51
5.1 OBTENÇÃO DOS AGRUPAMENTOS	51
5.2 COMPONENTES PRINCIPAIS	52
5.3 PREÇOS OBTIDOS NO MODELO PROPOSTO	55
5.4 CÁLCULO DO PREÇO DE NOVOS IMÓVEIS	59
5.5 COMPARAÇÕES	60
6 CONSIDERAÇÕES FINAIS	66
6.1 CONCLUSÕES	66
6.2 SUGESTÕES PARA TRABALHOS FUTUROS.....	67
REFERÊNCIAS	68
ANEXOS	69

LISTA DE FIGURAS

FIGURA 2.1 - LOCALIZAÇÃO DA CIDADE DE CAMPO MOURÃO – PR	6
FIGURA 2.2 - MUNICÍPIOS VIZINHOS À CAMPO MOURÃO.....	7
FIGURA 3.1 - O NEURÔNIO BIOLÓGICO	25
FIGURA 3.2 - O NEURÔNIO ARTIFICIAL.....	25
FIGURA 3.3 - FUNÇÕES DE ATIVAÇÃO DOS NEURÔNIOS: (A) DEGRAU; (B) IDENTIDADE; (C) SIGMOIDAL	26
FIGURA 3.4 - ARQUITETURA DE CAMADA ÚNICA.....	28
FIGURA 3.5 - ARQUITETURA DE MÚLTIPLAS CAMADAS.....	28
FIGURA 3.6 - ARQUITETURA RECORRENTE	28
FIGURA 3.7 - ESQUEMA ALGORITMO <i>BACK-PROPAGATION</i>	31
FIGURA 3.8 - ALGUNS PARÂMETROS DA REDE DE KOHONEN	38
FIGURA 3.9 - ANÁLISE DE COMPONENTES PRINCIPAIS	44
FIGURA 4.1 - FLUXOGRAMA EXPLICATIVO DO SISTEMA AIRNA.....	50
FIGURA 5.1 - CÁLCULO DE PREÇOS DE NOVOS IMÓVEIS.....	60

LISTA DE TABELAS

TABELA 5.1 - COMPONENTES PRINCIPAIS PARA OS APARTAMENTOS	52
TABELA 5.2 - COMPONENTES PRINCIPAIS PARA AS CASAS	52
TABELA 5.3 - COMPONENTES PRINCIPAIS PARA OS TERRENOS	55
TABELA 5.4 - MELHOR RESULTADO PARA OS APARTAMENTOS	56
TABELA 5.5. MELHOR RESULTADO PARA AS CASAS	57
TABELA 5.6. MELHOR RESULTADO PARA OS TERRENOS	58
TABELA 5.7 - COMPARAÇÃO REALIZADA PARA APARTAMENTOS	61
TABELA 5.8 - COMPARAÇÃO REALIZADA PARA CASAS	62
TABELA 5.9 - COMPARAÇÃO REALIZADA PARA TERRENOS.....	63

RESUMO

Neste trabalho é apresentada uma proposta para a avaliação de imóveis urbanos (apartamentos, casas e terrenos) através de técnicas de Análise Multivariada e Redes Neurais Artificiais. Os dados utilizados são os mesmos encontrados nos trabalhos de Alves (2005) e Braulio (2005), os quais utilizaram propostas baseadas apenas em Métodos Estatísticos Multivariados. A proposta deste trabalho consiste em: efetuar a padronização dos dados, separá-los em grupos por meio de um Mapa Auto-organizável de Kohonen, aplicar a Análise de Componentes Principais, com 98% de explicação da variabilidade dos dados, para evitar o problema clássico da multicolinearidade e, só então, aplicar em cada grupo uma Rede Neural de Múltiplas Camadas, para determinar os preços de venda dos imóveis e, posteriormente, comparar os resultados obtidos com os encontrados nos referidos trabalhos. Nesta comparação foi possível observar que as Redes Neurais mostraram um melhor desempenho na predição de preços dos imóveis, chegando a ser 299,17% superior na predição de preços dos apartamentos em relação a Alves (2005), sendo este o melhor resultado, e 14,39% superior na predição de preços em relação à Braulio (2005), também de apartamentos, sendo este o pior resultado.

Palavras-chave: Avaliação de Imóveis, Redes Neurais Artificiais *Feed-forward*, Algoritmo *Back-propagation*, Mapas Auto-organizáveis de Kohonen.

ABSTRACT

In this paper presents a proposal for the evaluation of urban property (apartments, houses and lands) combining techniques of Multivariate Statistical Analysis and Artificial Neural Networks. The data are the same ones found in Alves (2005) and Braulio (2005), which are based only on Multivariate Statistical Methods. The proposal of this paper consists in: to normalize the data, separate them in groups using Kohonen Self-organizing Maps, apply the Principal Component Analysis, using 98% of variability explication, to prevent the classic problem of the multicollinearity and, then, apply in each formed group a Multilayer Neural Network, to determine the sale prices of the property and, later, to compare the results gotten with the found ones in the related papers. In this comparison it could be seen that the Neural Networks had shown a better performance in the price prediction of the properties, more specifically 299.17% superior in the prediction of the apartments prices relative to Alves (2005), the best result, and 14.39% superior in the prediction of prices relative to Braulio (2005), apartments too, in the worse one result.

Key-words: Property Evaluation, Feed-forward Artificial Neural Networks, Back-propagation Algorithm, Kohonen Self-organizing Maps.

1 INTRODUÇÃO

Seres humanos podem, facilmente, baseados em exemplos e, muitas vezes, através de treinamento, com a ajuda de um professor, aprender a reconhecer um símbolo (ou caractere), diferenciar um animal de outro, aprender a eleger a melhor alternativa de um dado conjunto. Quanto mais exemplos forem "aprendidos", maior será a capacidade de distinção (ou decisão) adquirida. Entretanto, se a quantidade de informações for demasiadamente grande, a tendência é que o tempo para se tomar a decisão correta (ou seja, fazer a distinção entre os mesmos), torne-se impraticável ou, ainda, a decisão seja incorreta.

A criação de um modelo matemático que simule os neurônios humanos (e, conseqüentemente, a sua forma de aprendizado) pode auxiliar no processo de automação dessa tarefa, através da criação de padrões, permitindo diferenciar um grupo de outro, escolhendo uma dentre várias alternativas. Outra ferramenta importante para resolver esse tipo de problema é o uso de Análise Multivariada de Dados para tentar diminuir o número de variáveis, garantindo que variáveis irrelevantes não exerçam influência sobre o resultado final e, também, evitando o problema clássico da multicolinearidade dos dados.

1.1 PROBLEMA

A avaliação de imóveis, sejam urbanos ou rurais, faz-se presente na maioria dos negócios, pendências entre pessoas, empresas ou ambos. Geralmente é necessário avaliar imóveis para a compra e venda dos mesmos, na determinação do preço real de empresas, em atendimento à legislação, na partilha de heranças, divórcios, no lançamento de impostos, nas hipotecas imobiliárias, na demarcação de terras, nas indenizações, enfim, em um número expressivo de ações inerentes aos relacionamentos humanos, onde o valor de um bem assume importância crucial.

Toda vez que uma empresa necessita tomar dinheiro emprestado para investimento em qualquer segmento com o intuito de aumentar os bens ou obter capital de giro, recorre a bancos de desenvolvimento, os quais atendem a essas solicitações desde que o solicitante apresente garantias reais, isto é, tenha bens com valores iguais ou superiores ao valor do empréstimo. Entende-se por garantias reais, imóveis, máquinas e equipamentos, entre outros. Sendo assim, toda vez que se toma capital emprestado em bancos de desenvolvimento, há a necessidade de emissão de um laudo de avaliação do imóvel, máquina ou equipamento, para

que possam denotar garantias reais. Neste trabalho é estudado o caso de avaliação do valor de imóveis urbanos: casas, apartamentos e terrenos.

Apesar do conceito de valor ser de difícil definição, sujeito e suscetível às mudanças filosóficas, é importante que sejam adotados critérios para que haja justiça na determinação do valor de bens. Desse modo, avaliar imóveis pode ser definido como uma sequência de operações que têm como resultado a “formação de juízo” sobre o valor de um imóvel ou algum direito sobre ele.

A norma brasileira NBR5676 (ABNT, 2004) trabalha com o conceito de que o valor é aquele fornecido para um dado instante, não importando qual a finalidade da avaliação. Esse valor corresponde ao preço que se definiria, para um determinado imóvel, em um mercado de concorrência perfeito, o qual está sempre sujeito às seguintes premissas:

- a) Homogeneidade dos bens levados a mercado;
- b) Número elevado de compradores e vendedores;
- c) Ausência de influência externa;
- d) Conhecimento pleno e absoluto sobre o mercado, sobre os bens e das tendências de avaliação por parte dos compradores e vendedores;
- e) Vendedores e compradores oferecendo liquidez com liberdade plena de entrada e saída do mercado.

A partir destas considerações, pode-se afirmar que a avaliação passa a ser uma determinação técnica do valor ou de um direito sobre o imóvel, fato garantido pela liquidez com liberdade e a ausência de influência externa.

Para uma correta avaliação dos imóveis, dentre outros fatores, deve-se levar em conta que o valor de um bem está diretamente ligado à sua rentabilidade, sua possível utilização ou a sua raridade. Os imóveis urbanos podem ser definidos como bens que não são móveis, localizados nas cidades, geralmente classificados como glebas urbanizáveis, áreas ou lotes e terrenos com benfeitorias (casas, prédios residenciais, prédios comerciais, galpões e outros). A NBR 5676-ABNT (1989) define como sendo uma gleba urbanizável, “uma grande extensão de terreno passível de receber obras e infra-estruturas urbanas, por sua localização, seus aspectos físicos, sua destinação legal e pela existência de um mercado comprador”.

O que ocorre é que, em grande parte dos mercados imobiliários, a avaliação dos imóveis urbanos é feita subjetivamente, com base na experiência ou na “esperteza” dos

vendedores (para os casos onde o imóvel é vendido por um preço superior ao preço real) ou, ainda, com base nas necessidades financeiras da empresa imobiliária ou do proprietário do imóvel (para os casos onde o imóvel é vendido por um preço inferior ao preço real). Em ambas as situações, a ausência de uma base científica na avaliação dos imóveis pode fazer com que um dos lados tenha uma perda considerável no âmbito financeiro.

Assim, neste trabalho desenvolveu-se uma rotina em MATLAB 7.1® que, a partir das características particulares dos imóveis, estima seu valor de forma objetiva, ou seja, apenas em função de suas características (variáveis, atributos) sem qualquer influência de opiniões pessoais ou experiências humanas de vendedores.

A definição das variáveis a serem utilizadas é um dos pontos mais importantes para que a avaliação tenha o máximo de informações e de eficácia. Para a avaliação de imóveis urbanos, por exemplo, pode-se levar em conta as seguintes variáveis: a dimensão de testada, a profundidade, a área total, a localização, o uso do solo, as posturas municipais, o zoneamento urbano, as distâncias a pólos que os valorizem ou os desvalorizem, a taxa de ocupação, a topografia, a suscetibilidade a enchentes ou a danos ambientais, o padrão de construções na vizinhança, a infra-estrutura urbana e a paisagem visual a partir do imóvel. Tais variáveis devem ser “traduzidas” para a linguagem matemática de forma que conceitos matemáticos possam determinar o valor do imóvel.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Quando se busca a solução de um problema, obviamente, diferentes abordagens podem levar a resultados distintos, algumas vezes muito semelhantes, mas raramente iguais em todos os aspectos.

Assim, o presente trabalho visa dar tratamento científico à avaliação de imóveis urbanos e comparar a abordagem via Redes Neurais, desenvolvida neste trabalho, com o auxílio do *Software* MATLAB 7.1 com as abordagens via Métodos Estatísticos Multivariados propostas em Alves (2005) e Braulio (2005).

1.2.2 Objetivos Específicos

O objetivo geral pode ser atingido desde que se alcancem os seguintes objetivos específicos:

- a) Utilizar técnicas Multivariadas para a padronização dos dados, com o intuito de melhorar o desempenho dos algoritmos a serem aplicados;
- b) Utilizar as Redes de Kohonen para formação de grupos de imóveis, com o intuito de agrupá-los para a determinação dos preços posteriormente;
- c) Utilizar o método da Análise de Componentes Principais, como auxiliar para diminuição do número de variáveis, com a conseqüente eliminação das variáveis estatisticamente não significativas para os resultados finais bem como evitar a multicolinearidade dos dados;
- d) Aplicar em cada grupo uma Rede Neural Multicamada, para determinar os preços dos imóveis;
- e) Interpretar os resultados obtidos e compará-los com os resultados obtidos nos trabalhos de Braulio (2005) e Alves (2005).

1.3 JUSTIFICATIVA

A necessidade de estabelecer valor pode ser considerada uma das atividades mais freqüentes do homem moderno, sendo os imóveis um de seus bens mais valorizados. Então, a avaliação de imóveis é importante nas operações de compra e venda, locações, garantias de financiamentos, apólices de seguro, instalações comerciais e industriais, cobrança de impostos, ou seja, atividades cotidianas inerentes ao relacionamento humano.

O valor de um imóvel pode ser subjetivo dependendo das circunstâncias e modo de avaliação. Por outro lado, o valor de um imóvel é dado pela soma dos valores da edificação e pelo valor do terreno, que depende da sua localização. No entanto, o valor de um bem não pode ser confundido com o preço do bem, que representa a quantidade de dinheiro paga pelo mesmo, o que depende da lei da oferta e da procura.

1.4 ESTRUTURA DO TRABALHO

Esta pesquisa está estruturada em seis capítulos, visando uma melhor divisão dos assuntos e, conseqüentemente, um entendimento mais fácil por parte do leitor. O problema, os objetivos, a justificativa e a estrutura estão neste primeiro capítulo. O segundo capítulo traz uma descrição detalhada do problema, a separação das variáveis e a descrição da localidade envolvida. O terceiro capítulo traz os conceitos necessários ao entendimento do trabalho, tratando basicamente de conceitos de Estatística Multivariada e Redes Neurais Artificiais. O quarto capítulo traz a descrição de como as técnicas apresentadas no terceiro capítulo foram aplicadas, as definições de parâmetros, explicação do procedimento desenvolvido. O quinto capítulo traz os resultados obtidos, bem como, as análises inerentes a estes resultados e às comparações propostas durante o trabalho. Finalmente, o sexto capítulo traz as considerações finais sobre o trabalho e as sugestões para trabalhos futuros com os dados contidos neste trabalho.

2 DESCRIÇÃO DO PROBLEMA

2.1 A ÁREA DO ESTUDO

A área de estudo foi a cidade de Campo Mourão, situada no Noroeste do Estado do Paraná, distante 87 km de Maringá, 323 km de Foz do Iguaçu, 477 km de Curitiba e 659 km de São Paulo e se constitui no maior entroncamento rodoviário do Sul do Brasil. Sendo a sede da Microrregião 12 (divisão administrativa estadual), Campo Mourão agrega 25 municípios com economia baseada inicialmente no setor primário e hoje realiza investimentos na área industrial, já em avançado estágio de implementação do setor secundário e desenvolvimento do terciário.

A fertilidade da terra permite uma grande produtividade no campo. As principais culturas são: soja, trigo, milho, algodão e aveia. Paralelamente à agricultura, destaca-se o parque de revendas e assistência técnica de equipamentos e insumos. Campo Mourão é sede da maior cooperativa singular da América Latina, a Coamo - Cooperativa Agropecuária Mourãoense, hoje Coamo Agroindustrial Cooperativa. As Coordenadas geográficas do Município são 24°02'38" de Latitude Sul e 52°22'40" de Longitude Oeste do Meridiano de *Greenwich*, a uma altitude média de 630 metros sobre o nível do mar. A seguir, na figura 2.1, tem-se a localização da cidade no mapa do Estado do Paraná.

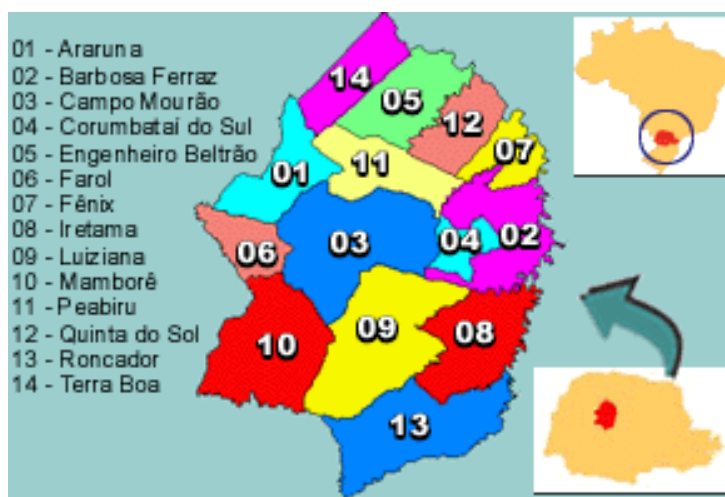
FIGURA 2.1 - LOCALIZAÇÃO DA CIDADE DE CAMPO MOURÃO – PR



FONTE: Prefeitura Municipal de Campo Mourão, 2005.

Campo Mourão limita-se com os seguintes municípios: Peabiru, ao norte; Barbosa Ferraz, a nordeste; Luiziana, ao sul; Corumbataí do Sul, a leste; Farol e Mamborê, a oeste e Araruna, a noroeste. A seguir, na figura 2.2, tem-se o mapa dos municípios vizinhos de Campo Mourão.

FIGURA 2.2 - MUNICÍPIOS VIZINHOS À CAMPO MOURÃO



FONTE: Prefeitura Municipal de Campo Mourão, 2005.

O Município de Campo Mourão pertence à bacia hidrográfica do Rio Ivaí, sendo seu rio mais importante o Rio Mourão, que atravessa o Município de sul a norte. Outros rios, importantes, por serem condicionantes físico-naturais à expansão urbana de Campo Mourão, são o Rio km 119 e Rio do Campo, este último, onde a SANEPAR coleta 80% da água que abastece o município.

A seguir têm-se outras informações: Área da unidade territorial: 766,44 km²; População estimada em 2004: 81.259 habitantes; Pessoas Residentes na Área Urbana: 74.754 habitantes; Domicílios particulares permanentes em 2004: 22.829 domicílios; atividades imobiliárias (aluguéis e serviços prestados às empresas): 36 empresas.

2.1.2 Limitações da Pesquisa

Um fator que pôde, visivelmente, ser considerado como limitante da pesquisa foi a falta de uma amostra maior para os três tipos de imóveis avaliados, sendo que o caso mais crítico certamente foi dos terrenos, com apenas 24 observações fornecidas pela maior imobiliária da cidade em questão, a TAPOWIC.

Para uma maior segurança e confiabilidade no tratamento matemático e/ou estatístico dos dados, é recomendado que o número de observações seja, no mínimo, o triplo do número de variáveis destas observações. Isso significa dizer que se cada observação tem 30 variáveis, por exemplo, o número necessário de observações para obter essa maior confiabilidade seria de, no mínimo, 90 observações.

Outra limitação do trabalho é o espaço de tempo. Como um dos objetivos do trabalho é comparar resultados obtidos neste trabalho com os obtidos por outra metodologia, os valores não foram atualizados e, conseqüentemente, não representam a realidade de preços atuais de mercado.

2.1.3 Levantamento dos Dados

Os dados utilizados nesta pesquisa são os mesmos da pesquisa realizada por Silvia Neide Braulio (2005), obtidos e organizados de forma cautelosa, sendo que dos dados depende o sucesso da Análise Matemática ou Estatística.

No planejamento realizado, pela referida pesquisadora, antes da coleta dos dados foi contemplado o espaço físico, população a estudar e o número de imóveis a serem pesquisados.

Vale observar que, no mercado de imóveis, é freqüente a entrada de dados novos e, por isso, deve-se fazer um novo levantamento a cada nova avaliação para garantir a representação dos novos dados na amostra (ALVES, 2005).

2.1.3.1 As variáveis utilizadas

As variáveis independentes (explicativas), dos tipos quantitativo e qualitativo, representando as características do imóvel, são detalhadas a seguir.

A variável dependente (resposta) é o preço, que representa o valor de venda do imóvel em reais, moeda corrente no Brasil. As variáveis originais e as independentes estão relacionadas, classificadas e descritas nos quadros 2.1 a 2.3 para os tipos de imóveis, apartamentos, casas e terrenos, respectivamente.

Uma consideração a ser feita é a de que a avaliação dos níveis como, por exemplo, de revestimento do prédio para apartamentos, podem ter interpretações diferentes dependendo da região de avaliação do imóvel.

Assim, no caso de uma implementação de um programa que calcule os preços de imóveis, não seria aconselhável utilizar dados de imóveis de uma região, por exemplo região sul do Paraná, para treinar a rede para determinar o preço de um imóvel de outra região, como por exemplo, região norte do Paraná. Uma outra saída seria obter dados de várias regiões mas, possivelmente, a generalização da rede não seria tão eficaz quanto a abordagem regionalizada.

QUADRO 2.1 - VARIÁVEIS INDEPENDENTES PARA APARTAMENTOS

continua

Variáveis	Categorias	Descrição
Revestimento do prédio	(1) reboco/emboço (2) tinta plástica (3) pastilhas (4) mármore	Identifica o revestimento do prédio.
Andar	(1) 1º andar (2) 2º andar (3) 3º andar ou mais (4) cobertura	Identifica o andar que o apartamento está localizado. Sabe-se que dependendo do andar que se localiza o apartamento ele é mais ou menos valorizado.
Dependência de empregado	(0) sem (1) com	Identifica a existência ou não de dependência de empregados.
Estado de conservação	(1) péssimo (2) regular (3) bom (4) ótimo	Identifica o nível de conservação do imóvel.
Suíte	(0) sem (1) com	Identifica a presença ou não de suíte, atribuindo o valor 1 mesmo quando há presença de mais de uma suíte.
Idade aparente (em anos)	(1) 20 – ∞ (2) 15 – 20 (3) 11 – 14 (4) 6 – 10 (5) 2 – 5 (6) 0 – 1	Idade aparente: idade aparente do edifício. Por ser uma variável contínua, a idade do imóvel dividiu-se em períodos.
Idade real (em anos)	(1) 20 – ∞ (2) 15 – 20 (3) 11 – 14 (4) 6 – 10 (5) 2 – 5 (6) 0 – 1	Idade real: idade cronológica do edifício reflete o estágio tecnológico.
Proximidade	(1) mais de 800m (2) 500m – 800m (3) até 500m	Identifica a quanto o imóvel se localiza próximo de escolas, supermercados, hospitais e do centro comercial.

conclusão

Variáveis	Categorias	Descrição
Lavanderia	(0) sem (1) com	Identifica a existência ou não de lavanderia.
Posição do apartamento	(1) fundo (2) lateral (3) frente	Identifica a posição do apartamento em relação ao prédio (frente, lateral ou fundo).
Padrão de acabamento	(1) baixo (2) normal (3) alto	Identifica os vários níveis de acabamento.
Sala	Unidades	Indica o número de salas existentes no apartamento.
Pavimento	Unidades	Indica o número de pavimentos do prédio.
Garagem	Unidades	Quantifica o número de vagas para carro disponível para cada apartamento.
Dormitório	Unidades	Quantifica o número de dormitórios.
Elevador	Unidades	Identifica a quantidade de elevadores no prédio.
Área privativa	m^2	Corresponde à superfície ou área do apartamento expressa em metros quadrados, obtida do registro de imóveis.
Peças	Unidades	Quantifica as peças constituintes do imóvel.
Banheiro	Unidades	Identifica o número de banheiro social.
Pólo de influência	(-1) pólo desvalorizante (0) inexistente (1) pólo valorizante	Indica se o móvel localiza-se próximo a locais que influenciam no seu valor.

Fonte: Imobiliária Tapowik, 2005.

QUADRO 2.2 - VARIÁVEIS INDEPENDENTES PARA CASAS RESIDENCIAIS

continua

Variáveis	Categorias	Descrição
Localização	(1) inferior (2) razoável (3) bom (4) ótimo (5) centro	Naturalmente um local é “melhor” ou “pior” do que um outro em função de diversas características, entre as quais sua infra-estrutura urbana.
Dependência de empregado	(0) inexistente (0,5) incompleta (1) completa	Identifica a existência ou não de dependência de empregado, completa ou incompleta.
Nível de conservação	(1) péssimo (2) regular (3) bom (4) ótimo	Identifica o nível de conservação do imóvel.
Suíte	(0) sem suíte (1) com suíte	Identifica a presença ou não de suíte, atribuindo o valor 1 mesmo quando há presença de mais que uma suíte.

conclusão

Variáveis	Categorias	Descrição
Idade aparente (em anos)	(1) 20 – ∞ (2) 15 – 20 (3) 11 – 14 (4) 6 – 10 (5) 2 – 5 (6) 0 – 1	Por ser uma variável contínua a idade do imóvel dividiu-se em períodos.
Garagem	(0) sem (1) com	Identifica a presença de garagem, onde é atribuído o valor mesmo quando a mais que uma vaga.
Distância de supermercados	(1) mais de 800m (2) 500m – 800m (3) até 500m	Identifica a proximidade do imóvel de grandes mercados.
Presença de lavanderia	(0) sem (1) com	Identifica a existência ou não de lavanderia.
Edícula	(0) sem (1) com	Identifica a presença (1) ou não (0) de edícula.
Padrão de acabamento	(1) baixo (2) normal (3) alto	Identifica os vários níveis de acabamento.
Piscina	(0) sem (1) com	Identifica a existência ou não de piscina.
Cobertura	(1) madeira+eternit (2) madeira+telha (3) laje+eternit (4) laje+telha	Identifica o tipo de cobertura do imóvel.
Estrutura	(1) mista (2) madeira (3) alvenaria (4) sobrado	Identifica o material de construção do imóvel.
Dormitório	Unidades	Quantifica o número de dormitórios.
Área do terreno	m^2	Identifica a área do terreno.
Área construída	m^2	Identifica a área total construída.
Peças	Unidades	Quantifica as peças constituintes do imóvel.
Banheiro	Unidades	Identifica o número de banheiro social.

Fonte: Imobiliária Tapowik, 2005.

QUADRO 2.3 - VARIÁVEIS INDEPENDENTES PARA TERRENOS

continua

Variáveis	Categorias	Descrição
Localização	(1) péssimo (2) ruim (3) regular (4) aceitável (5) bom (6) centro	Variável que qualifica a localização do imóvel.

conclusão

Variáveis	Categorias	Descrição
Pólo de influência	(-1) pólo desvalorizante (0) inexistente (1) pólo valorizante	Indica se o imóvel localiza-se próximo a locais que influenciam no seu valor
Plano	(1) acima (2) abaixo (3) mesmo nível da rua	Identifica se o terreno está acima, abaixo ou ao nível da rua.
Inclinado	(1) muito inclinado (2) pouco inclinado (3) requer aterro	Indica o nível de inclinação do terreno.
Pavimentação	(0) sem pavimentação (1) com pavimentação	Identifica a presença ou não de pavimentação na rua ou avenida onde está inserido o terreno.
Proteção	(0) sem proteção (1) com proteção	Indica se o terreno possui ou não proteção (muro ou cerca).
Posição	(1) meio (2) esquina	Identifica a posição do terreno na quadra (meio ou esquina).
Frente	(1) até 10m (2) 10m – 15m (3) mais de 15m	Identifica a largura do terreno. Sabendo que um terreno de frente com maior metragem possui uma melhor valorização.
Ponto Comercial	(0) moradia (1) bairro (2) periferia central (3) centro	Sabendo que os terrenos localizados em zona de comércio ou de moradia, o terreno é mais ou menos valorizado. Esta variável identifica os vários níveis de localização.
Área do terreno	m^2	Quantifica a área do terreno.

Fonte: Imobiliária Tapowik, 2005.

Vale salientar que muitas variáveis são extremamente subjetivas, como por exemplo, a “idade aparente” de um imóvel carece de critérios para especificação de seus valores exibidos nas tabelas em anexo a este trabalho. Contudo, como o cunho deste trabalho é comparativo, optou-se pela manutenção dos critérios originais dos trabalhos de Alves (2005) e Braulio (2005).

2.1.3.2 Amostra

A amostra foi constituída por 119 imóveis. Sendo 44 apartamentos, 51 casas e 24 terrenos localizados na área urbana da Cidade de Campo Mourão – PR, dos quais 80 estão localizados na área central.

2.2 METODOLOGIA PARA O DESENVOLVIMENTO DA PESQUISA

A metodologia aqui proposta procura determinar, inicialmente, grupos homogêneos

(*clusters*) de apartamentos, casas e terrenos através das Redes de Kohonen, técnica de Redes Neurais Artificiais utilizada para agrupamentos, aplicada à amostra considerada para confrontar resultados com os trabalhos de Alves (2005) e Braulio (2005).

Então, a partir dos grupos formados, aplica-se o método da Análise de Componentes Principais, procurando eliminar o problema da multicolinearidade e, possivelmente reduzir variáveis. Dessa forma, apenas as variáveis altamente explicativas da variabilidade dos dados permanecem no modelo.

Após a aplicação da técnica da Análise de Componentes Principais, em cada uma das grupos, é aplicada uma Rede Neural Multicamada com a utilização do algoritmo *back-propagation* com taxa de momento padrão do MATLAB (0,9) e taxa de aprendizagem adaptativa de acordo com o erro a cada iteração. Se há queda no desempenho da rede, a taxa de aprendizagem é incrementada em valor; se há aumento excessivo provocando *overfitting*, a taxa de aprendizagem é decrementada em valor.

Todas estas etapas são desenvolvidas por meio de um sistema, desenvolvido em linguagem MATLAB, ao qual foi dado o nome de AIRNA – Avaliação de Imóveis via Redes Neurais Artificiais.

3 REVISÃO DE LITERATURA

O objetivo deste capítulo é apresentar da forma mais clara e objetiva possível os principais conceitos necessários ao entendimento do modelo matemático proposto, bem como a engenharia de avaliação, mais especificamente, a avaliação de imóveis.

3.1 ENGENHARIA DE AVALIAÇÃO

As primeiras publicações sobre avaliação de imóveis no Brasil surgiram por volta de 1918; em 1923 já houve uma melhoria nos métodos de avaliação de terrenos, que a partir de 1929, começaram a serem sistematicamente utilizados. A partir daí a engenharia de avaliação no Brasil começou a ganhar espaço, continua ganhando e a apresentando técnicas cada vez mais evoluídas de avaliação. Atualmente, o número de publicações e estudos na área aumentou consideravelmente fazendo com que a avaliação de imóveis tenha um suporte científico consistente, com completa ausência de subjetividade no momento da avaliação.

Para Dantas (*apud* BRAULIO, 2005), a Engenharia de Avaliações é um dos campos da engenharia que reúne uma gama de conhecimentos relativos à engenharia e arquitetura, bem como diversas áreas do conhecimento: ciências sociais, exatas e da natureza. Tem por objetivo determinar tecnicamente, ou seja, sem subjetividade alguma, o valor de um bem, de seus direitos, frutos e custos de reprodução.

A Engenharia de Avaliações pode ser praticada por engenheiros de diversas áreas, obviamente, cada um na sua respectiva habilitação e arquitetos, conforme normas e regulamentos do CREA, CONFEA, ABNT, leis municipais, estaduais e federais.

3.1.1 Histórico Sobre as Normas Técnicas

A ABNT - Associação Brasileira de Normas Técnicas é o Fórum Nacional de Normatização. As Normas Brasileiras, cujos conteúdos são de responsabilidade dos Comitês Brasileiros (ABNT/CB) e dos Organismos de Normalização Setorial (ABNT/ONS) e das Comissões de Estudos Especiais Temporários (ABNT/CEET), são elaboradas por Comissões de Estudo (CE), formadas por representantes dos setores envolvidos, delas fazendo parte: produtores, consumidores e neutros (universidades, laboratórios e outros) (ABNT NBR 14653-2, 2004).

Segundo Dantas (*apud* ALVES, 2005) em meados de 1950 surgiram as primeiras normas de avaliação de imóveis organizadas por entidades públicas e institutos. Devido à ocorrência de grande quantidade de desapropriações na cidade de São Paulo, ocasionado pela expansão da cidade e construção do metrô na década de 1960, as normas ganharam maior relevância. Porém, o primeiro anteprojeto de normas da ABNT na Engenharia de Avaliação data de 1957. Em 1977, estudos feitos por comissões de profissionais dedicados às perícias e avaliações judiciais, em essência, deram origem à primeira Norma Brasileira para Avaliação de Imóveis Urbanos, a NB-502/77 da ABNT.

Depois de passar por uma revisão em 1989, a Norma Brasileira para Avaliação de Imóveis Urbanos foi registrada no INMETRO como NBR 5676. Os níveis de precisão foram transformados em níveis de rigor. Segue-se a ela a Norma para Avaliação de Servidões. Alguns institutos, com base na NBR 5676, produziram, paralelamente, normas específicas de forma mais detalhada observando as características de cada região. Atualmente existe uma proposta levada à ABNT para sintetizar o tema de Engenharia de Avaliações em uma norma única. Trata-se da Norma para Avaliação de Bens, formada por uma parte principal, contendo os conceitos, métodos e definições comuns a todos os bens e, nos apêndices, a parte específica para cada tipologia de bem a avaliar.

3.1.1.1 Valor

Desde sua origem, o homem busca estabelecer preços para os bens, tais que satisfaçam sua noção de “justiça” numa transação e de maneira que se efetivem trocas, sejam elas diretas como o escambo (permuta) ou indiretas como a moeda.

O conceito de valor de um bem, de modo geral, é intuitivo e subjetivo, quer seja vendedor ou comprador deste bem, podendo variar entre os participantes de um mercado. No entanto, o preço é uma característica que representa a quantidade de dinheiro (no caso da troca indireta) paga pelo bem.

Para Fiker (1997), valor, em economia política, é a relação entre o grau das necessidades econômicas do homem e a quantidade de bens disponíveis para satisfazê-las. O fator necessidade envolve aspectos objetivos e subjetivos. Como aspecto objetivo pode-se citar, por exemplo, comprar alimentos para saciar a fome. Já o aspecto subjetivo está relacionado ao desejo de obter o bem que, entretanto, ainda é uma necessidade, provocada por um aspecto subjetivo, o desejo.

Segundo Braulio (2005), muitas medidas de valor podem estar relacionadas a um bem, dentre elas o custo de produção, ao qual são agregados outros custos como matéria-prima, estocagem e comercialização desde o produtor até o produto final onde será formado o preço e o valor de mercado, não havendo necessariamente uma relação matemática entre eles. No entanto, em mercados que se aproximam daquele de concorrência perfeita, os preços são estabelecidos pela lei da oferta e procura independentemente dos custos de produção. Em vista disso, no mercado considerado, o valor do bem poderá não apresentar nenhuma relação com os custos citados (podendo mesmo ser inferior).

Quando o mercado permanece estável por um longo período, o preço e a quantidade acabam sendo negociados através da oferta e procura. A partir daí toda a subjetividade e intuição que leva os participantes do mercado a tentar impor a sua opinião tornam-se, portanto, quantidades vendidas e ofertadas e seus respectivos preços. Logo, o preço estabelecido pelo mercado é considerado uma representação justa, embora subjetiva, do valor do bem analisado. As avaliações pelo valor de mercado podem ser consideradas instantâneas, ou seja, são válidas, apenas, por um intervalo curto de tempo.

Encontram-se várias definições e interpretações e que são suscetíveis a mudanças sobre os conceitos de valor, valor de mercado e preço. No entanto, torna-se importante determinar alguns critérios para sua aplicação prática. Assim, um trabalho de avaliação imobiliária constitui-se de uma série de operações e etapas até que se chegue a uma definição de valor. Dentre os diversos conceitos de valor, a Norma NB-502/89 (NBR-5676) da ABNT, de Avaliação de Imóveis Urbanos, define valor como sendo aquele fornecido para um bem em um dado instante, único, não importando qual a finalidade da avaliação. Esse valor corresponde ao valor real que se definiria em um mercado de concorrência perfeita caracterizado pelos seguintes silogismos:

- a) Igualdade dos bens levados ao mercado;
- b) Número elevado de compradores e vendedores;
- c) Sem influências externas;
- d) Conhecimento pleno e absoluto entre os participantes sobre o bem, o mercado e as tendências deste;
- e) Os participantes oferecendo liquidez com plena liberdade de entrada e saída do mercado.

Existem casos onde a necessidade de estimar um valor acontece a nível particular. Assim, as partes envolvidas vendedor e comprador do bem, chegam ao comum acordo da quantidade necessária (moedas) em um determinado instante, ambos com avaliação subjetiva na maioria dos casos. Porém, quando ocorre a necessidade da determinação do valor, de uma maneira mais ampla, isto é, sem ser a nível particular, portanto ampliado a outras pessoas além dos diretamente envolvidos na transação de ordem privada ou pública, procura-se uma perspectiva técnica, para garantir a imparcialidade na avaliação. Surge, então, a “Ciência da Avaliação”, ou seja, a Engenharia de Avaliações, que infere sobre o valor de um bem de forma fundamentada.

3.1.2 Avaliação e seus Métodos

A Norma NB-502/89 (NBR-5676) da ABNT, de avaliação de imóveis urbanos, define avaliação como a determinação técnica do valor de um imóvel ou de um direito sobre o imóvel. A avaliação de imóveis é a definição técnica do valor de mercado dos bens ou de direitos sobre eles. Esta definição é feita dentro de procedimentos técnicos para a realização das análises de valor.

Segundo Moreira (*apud* ALVES, 2005), avaliar é a arte de estimar valores apropriados e específicos, em que o conhecimento técnico e o bom-senso são condições fundamentais.

A seguir são apresentadas as descrições para quatro dos métodos de avaliação de bens descritos na norma NBR 14653-2:2004.

Vale ressaltar que existem mais alguns métodos, entretanto optou-se por citar somente esses quatro métodos que já são suficientes à pesquisa e desenvolvimento do modelo matemático.

3.1.2.1 Método comparativo direto de dados

Este método tem por objetivo identificar o valor de mercado de um bem por meio de tratamento técnico dos atributos dos elementos comparáveis, constituintes na amostra (ABNT NBR 14653-1, 2004).

3.1.2.2 Método involutivo

Identifica o valor de mercado do bem, alicerçado no seu aproveitamento eficiente, baseado em modelo de estudo de viabilidade técnico-econômica, mediante hipotético empreendimento compatível com as características do bem e com as condições do mercado no qual está inserido, considerando-se cenários viáveis para execução e comercialização do produto (ABNT NBR 14653-1, 2004).

3.1.2.3 Método evolutivo

Identifica o valor do bem pelo somatório dos valores de seus componentes. Caso a finalidade seja a identificação do valor de mercado, deve ser considerado o fator de comercialização (ABNT NBR 14653-1, 2004).

3.1.2.4 Método da capitalização da renda

Identifica o valor do bem, com base na capitalização presente da sua renda líquida prevista, considerando-se cenários viáveis (ABNT NBR 14653-1, 2004).

3.1.3 Diretrizes para a Aplicação dos Métodos

No planejamento de uma pesquisa, é necessário que haja uma amostra, suficientemente representativa, de dados de mercado de imóveis com características, tanto quanto possível, semelhantes às do imóvel a ser avaliado, usando-se toda a evidência disponível.

Isso sugere que esta etapa deve iniciar-se pela caracterização e delimitação do mercado a ser analisado, para tentar evitar heterogeneidade excessiva nos dados, com o auxílio de teorias e conceitos existentes ou hipóteses provenientes de experiências adquiridas pelo avaliador sobre a formação do valor.

Na estrutura da pesquisa são eleitas as variáveis independentes que, em princípio, são significativas para explicar a formação de valor e estabelecidas supostas relações entre si e com a variável dependente, no caso o valor do bem.

A estratégia de pesquisa refere-se, então, à abrangência da amostragem, às técnicas a serem utilizadas, seleção e abordagem das fontes de informação bem como a seleção do tipo

de análise (quantitativa ou qualitativa) e a elaboração dos instrumentos para a coleta de dados (fichas, planilhas etc.) (ABNT NBR 14653-2, 2004).

De acordo com a NBR 14653-2 (2004) para especificar corretamente a variável dependente é necessário observar o mercado: sua conduta, formas de expressão do preço e, também, a homogeneidade das unidades de medida. As variáveis independentes devem ser escolhidas com base em teorias existentes, conhecimentos adquiridos e outros atributos que se revelem importantes no decorrer dos trabalhos, pois algumas variáveis consideradas no planejamento podem, na verdade, ser pouco relevantes no resultado final e vice-versa. Para tentar evitar esse tipo de problema, surge a necessidade de utilizar-se de técnicas para a redução do número de variáveis não-significativas.

Para Fiker (1997) o melhor dos métodos de avaliação é o método comparativo por utilizar a comparação de dados do mercado relativos a outros imóveis, com características similares, na obtenção do valor do referido imóvel. Com base nessa afirmação, optou-se pela utilização do método comparativo neste trabalho.

Tabelas e também modelos de preços satisfatórios, tais como a Regressão Linear Múltipla, são utilizados para estimar valores. No entanto, ambos sofrem contestações sobre sua eficácia. As tabelas, que eram tradicionalmente utilizadas, são criticadas por serem imprecisas e de pouca confiabilidade. A Regressão Linear Múltipla tem demonstrado sérios problemas de multicolinearidade nas variáveis explicativas e também de inclusão de *outliers* na amostra.

Além disso, a colinearidade dentro dos dados pode tornar a Regressão Linear Múltipla um modelo inadequado para um mercado que requer respostas rápidas e de alta precisão, sendo aceitável apenas quando realizada por um profissional capacitado.

Em virtude disso, neste trabalho optou-se por utilizar a técnica de Redes Neurais, com o auxílio da técnica de Análise de Componentes Principais, para verificar se para esse tipo de problema há um desempenho melhor do que com métodos estatísticos multivariados.

3.1.4 Classificação dos Imóveis Urbanos

De acordo com a norma NBR-502/89 (NBR-5676) da ABNT, os imóveis são classificados em:

- a) Quanto ao uso, o imóvel urbano pode ser: residencial, comercial, industrial,

institucional e misto;

- b) Quanto ao tipo do imóvel, o imóvel urbano pode ser: terreno (lote ou gleba), apartamento, casa, escritório (sala ou andar corrido), loja, galpão, vaga de garagem, misto, hotéis, hospitais, cinemas e teatros, clubes e recreativos;
- c) Quanto ao agrupamento, os imóveis urbanos se agrupam da seguinte forma: loteamento, condomínio de casas, prédio de apartamentos, conjunto habitacional (casas, prédios ou mistos), conjunto de salas comerciais, prédio comercial, conjunto de prédios comerciais, conjunto de unidades comerciais, *shopping-centers* e complexo industrial. Este trabalho utilizou dados correspondentes apenas aos imóveis dos tipos: apartamentos, casas residenciais e terrenos.

3.1.5 O Mercado

Para Alves (2005) o mercado é o local onde são efetuadas as transações comerciais envolvendo troca de bens, tangíveis ou intangíveis, ou direitos sobre os mesmos. Neste sentido, o termo mercado refere-se àquele de concorrência perfeita e contendo, em geral, as características dos bens. Os participantes do mercado o fazem voluntariamente e têm conhecimento pleno das condições em vigor; nenhum participante, sozinho, é capaz de alterar as condições estabelecidas; cada transação é feita de maneira independente das demais; o número de ofertas e/ou transações é suficientemente grande, de maneira que a retirada de uma amostra não afeta o mercado.

3.1.5.1 Mercado imobiliário

Para Moscovitch (*apud* ALVES, 2005), o mercado imobiliário é a jurisdição de determinação dos preços de imóveis urbanos que, como quaisquer outras mercadorias, passam pela medida da oferta e da demanda, sem dúvida uma das leis econômicas mais conhecidas.

Tal mercado é formado por três secções:

1. a dos imóveis a serem vendidos;
2. a das partes que desejam vendê-los (vendedores);
3. a das partes interessadas em adquiri-los (compradores).

O mercado imobiliário pode ser subdividido ainda, em várias especialidades. As especialidades apartamentos, casas e terrenos, todos urbanos, foram as analisadas neste trabalho.

Cada mercado tem seu próprio comportamento e suas características específicas. No entanto, o mercado que possui, sem dúvida, o comportamento mais atípico entre os mercados de bens é o mercado de imóveis, devido às grandes diferenças entre os imóveis. Por sua localização fixa, qualquer alteração no ambiente provoca modificações no valor do imóvel. Como as influências não são análogas, as variações provocadas são claramente notáveis, causando progressivamente as diferenças.

Por outro lado, como todo bem econômico, a escassez relativa à lei da oferta e procura define o preço dos imóveis. Os governos e as economias globais são grandes influenciadores sobre o preço dos imóveis. E, por sua importância e significado social, as leis propiciam tratamentos especiais.

No mercado onde ocorrem as transações imobiliárias, identificam-se alguns fenômenos como o dinamismo da atividade imobiliária, fato que torna ainda mais difícil avaliar um imóvel, e o processo de estruturação interna das áreas urbanas. Existem, também, influências externas, que alteram continuamente os valores dos imóveis.

O estudo de todos estes fatores, que, cientificamente, são vistos como um vetor de características, constituem o processo de formação de valores, ou seja, determinação do preço ou de direitos sobre os imóveis.

Esses valores muitas vezes sofrem transformações como condições de mercado e também valores que são praticados, devido à falta de ordenação, utilização das normas, entre empreendedores, intermediários, poder público e também a própria população. Por se tratar de bens econômicos, todas as mudanças que causam maior ou menor disponibilidade refletem em modificações, alterações de valor.

Sabe-se que muitas alterações como a oferta de crédito, a inflação, a condução da economia, as políticas fiscais, o crescimento demográfico e a confiança no governo são importantes na flutuação de preços, por exemplo, a simples construção do campus de uma universidade conhecida em determinada região faz com que a procura por imóveis para alugar ou comprar aumente e, conseqüentemente, provoca aumento nos preços destes.

Mesmo com todo o dinamismo do mercado imobiliário, em certo momento, ocorrerá um fenômeno chamado de “equilíbrio instantâneo” que resultará num valor de mercado para

aquele imóvel naquele instante. Todas as variações ocorridas nas condições do mercado são incorporadas pelos imóveis, influenciando em seus valores, que oscilam no tempo e no espaço e que, em última análise, são resultados da oferta e demanda por este bem. E assim, outras mudanças na oferta ou na demanda causarão novo equilíbrio, em outros níveis de preço em outro momento.

Faz-se interessante observar que por existirem inúmeros fatores e influências, uma parte das variações dos valores imobiliários é considerada aleatória podendo-se pensar no preço final como sendo baseado em um valor mais provável que aumenta ou diminui por uma parcela imprevisível, variações do mercado devido aos fatores citados anteriormente, de acordo com as influências casuais.

Num mercado de concorrência perfeita a situação ideal é aquela onde a oferta e a procura encontram-se equilibradas sendo a relação entre imóvel, vendedor e comprador que são os formadores do mercado, especialmente importantes para a construção do preço. No entanto, pode ocorrer o mercado de concorrência imperfeita, ocasionando casos de monopólio e oligopólio.

Nos casos de monopólio o mercado torna-se comandado por um único vendedor, mas é um caso mais raro de ser encontrado. Entretanto é mais comum ser encontrado casos de oligopólio, onde a oferta é concentrada em um pequeno número de vendedores que também se preocupam com propagandas e qualidade dos imóveis, fazendo com que haja sempre uma alta nos preços dos imóveis. Ocorrem ainda os casos de monopsônio, existindo apenas um comprador e também oligopsônio com muitos compradores (Dantas, 1998).

Para Alves (2005) o mercado imobiliário não pode ser considerado um mercado de concorrência perfeita diante de muitas análises teóricas, pois para que pudesse ser considerado como tal deveria ser uma passagem livre na qual os bens são idênticos e que todos os participantes têm as informações perfeitas e que não sofrem nenhuma pressão agindo livremente.

No mercado imobiliário, ocorrem fatores que o dificultam: a falta de uniformidade dos imóveis e a falta de informações são exemplos disto. Não se pode esquecer que os altos custos impossibilitam grande parte da população a participar deste mercado de compra e venda, “sobrando” para esta parte de população apenas a locação e a financiamentos com juros muitas vezes absurdos, dados os altos valores dos imóveis. Tudo isso leva a crer que o mercado imobiliário é um mercado de concorrência imperfeita.

3.2 REDES NEURAIS ARTIFICIAIS

As Redes Neurais Artificiais (RNA's) são compostas de muitos elementos simples, inspirados pelo sistema nervoso biológico, que operam em paralelo. A função da rede é determinada pelas conexões entre os seus elementos e é possível treiná-la para que execute uma função particular ajustando-se os valores das conexões entre os elementos que a compõe (STEINER, 1995).

3.2.1 Um Breve Histórico

O estudo de Redes Neurais tem uma história de mais de seis décadas, mas suas aplicações se tornaram mais robustas somente nos últimos 25 anos. O campo de aplicação está se desenvolvendo rapidamente. Fausett (1994) descreve um breve histórico das Redes Neurais Artificiais:

- Em 1943, Mc Culloch e Pitts propuseram um modelo para uma célula nervosa, chamado neurônio formal ou neurônio artificial. Eles mostraram que um conjunto de neurônios era capaz de calcular certas funções lógicas particulares;
- Em 1949, Hebb descobriu o significado das conexões entre as sinapses, que são as conexões entre os neurônios, no processo de aprendizagem, e desenvolveu uma regra de aprendizagem básica, que ficou conhecida como “regra de Hebb”. Ele propôs que as mudanças nas forças das sinapses fossem proporcionais às ativações dos mesmos;
- Em 1959, Rosenblatt, numa junção das idéias de Hebb, McCulloch e Pitts, descreveu o primeiro modelo de Rede Neural, o Perceptron. Ajustando os neurônios em uma rede com uma topologia particular e variando os valores associados às conexões entre as sinapses, o Perceptron poderia aprender certas funções lógicas particulares;
- Em 1962, Widrow desenvolveu um tipo diferente de processador para Redes Neurais, denominado ADALINE, o qual dispunha de uma poderosa estratégia de aprendizado;
- Em 1969, este direcionamento da pesquisa em Redes Neurais foi abandonado quase que completamente, por força do trabalho de Minsky e Papert que expuseram as limitações do Perceptron;
- A década de 70 é denominada por Fausett (2004) como “os anos silenciosos”. Entretanto, em 1974, Werbos conseguiu o maior progresso em termos de Redes Neurais desde o Perceptron de Rosenblatt. Ele lançou as bases do algoritmo *back-*

propagation (retro-propagação do erro), que permitiu que Redes Neurais com múltiplas camadas apresentassem capacidade de aprendizado;

- Os primeiros resultados da retomada do desenvolvimento sobre Redes Neurais foram publicados em 1986 e 1987, onde ficou consagrada a técnica de treinamento por retro-propagação.

3.2.2 Características das Redes Neurais Artificiais

Para Steiner (1995) o sistema nervoso humano, o qual é constituído por aproximadamente de 200 bilhões de células, apresenta as seguintes principais características:

- É uma rede altamente interconectada;
- Apresenta paralelismo maciço, ou seja, muitos neurônios operando ao mesmo tempo;
- O processamento é distribuído de modo que a informação é não localizada, significando que um fato pode corresponder à atividade de certo número de neurônios;
- Admite tolerância a falhas, assim o prejuízo a poucos neurônios não afeta a operação do cérebro significativamente;
- A aprendizagem é exibida pelo ajustamento do efeito de acoplamento de 2 neurônios.

Apesar de o cérebro humano possuir um funcionamento muito complexo, ele é composto por elementos extremamente simples, os neurônios.

A estrutura de um neurônio biológico está apresentada na figura 3.1, um neurônio artificial na figura 3.2. A unidade básica de nosso cérebro, o neurônio, apresenta uma região onde informações são processadas (corpo celular), algumas entradas (os dendritos) e uma saída (o axônio), conforme figura 3.1, os impulsos elétricos recebidos nos dendritos são processados e o resultado deste processamento é colocado no axônio.

FIGURA 3.1 - O NEURÔNIO BIOLÓGICO

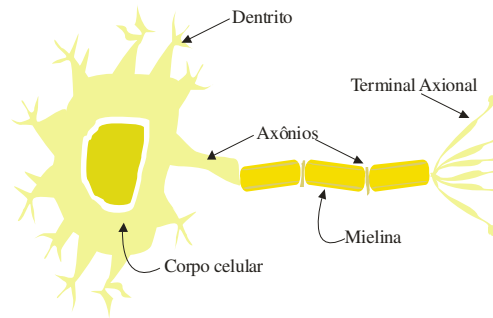
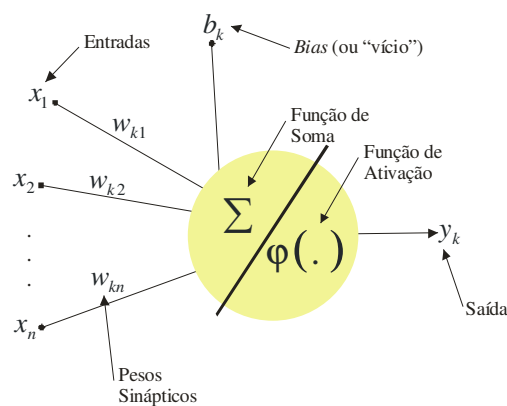


FIGURA 3.2 - O NEURÔNIO ARTIFICIAL



Na figura 3.2 os valores denotados por x_i são chamados de entradas do neurônio; o valor representado por y_k é a saída (resultado); e o valor b_k é chamado de *bias*, ou vício, da rede, trata-se de um peso adicional cujo valor de entrada é sempre igual a um.

3.2.2.1 Funções de ativação

Um neurônio pode ser representado matematicamente pelas expressões:

$$v_k = \sum_{j=1}^n w_{kj} x_j + b_k \quad (3.1)$$

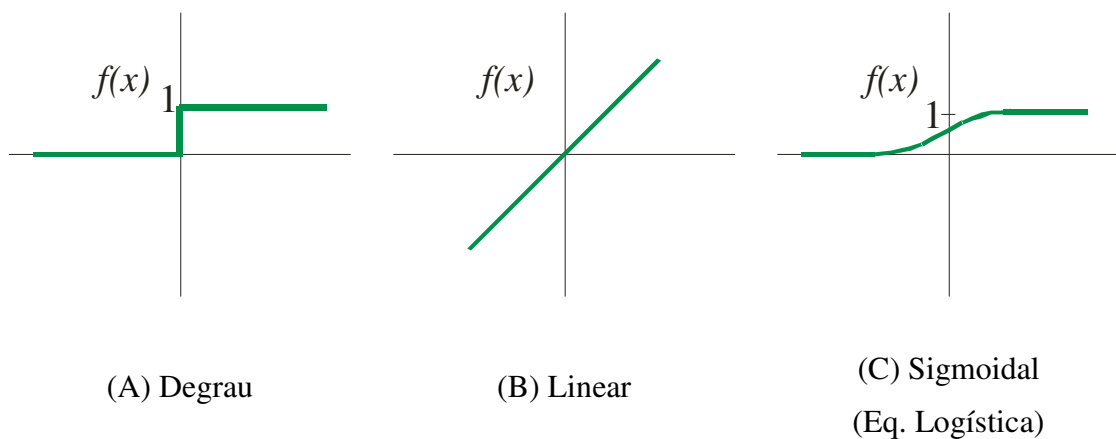
$$y_k = \phi(v_k) \quad (3.2)$$

A equação (3.1) determina o valor a ser lançado na função de ativação escolhida; na equação (3.2) a função ϕ é a chamada função de ativação do neurônio. É desejável que tal função seja contínua e diferenciável, em virtude da necessidade da utilização de sua derivada.

As funções de ativação podem ser de vários tipos, os três tipos mais utilizados, encontrados na figura 3.3, são: a função degrau, função linear e função sigmoidal.

Estes vários tipos de funções de transferência, dentre os quais se destacam três tipos de funções, apresentados na figura 3.3 a seguir, diferem apenas pelo que é normalmente chamado de “grau de exigência” da saída. Por exemplo, a função degrau tem saída binária, já as funções linear e sigmoidal são mais flexíveis, sendo a identidade a mais flexível.

FIGURA 3.3 - FUNÇÕES DE ATIVAÇÃO DOS NEURÔNIOS: (A) DEGRAU; (B) LINEAR; (C) SIGMOIDAL (EQ. LOGÍSTICA)



Normalmente, as funções não-lineares são as mais utilizadas. Outra prática comum é a de utilizar uma mesma função de ativação para todos os elementos de uma mesma camada em uma rede e, até mesmo, para toda a rede, entretanto, não é uma regra.

As equações (3.3), (3.4), (3.5) e (3.6) são relativas aos três tipos de função de ativação mais utilizados e podem ser encontrados, por exemplo, em Fausett (1994):

i) Função Degrau (ou *Threshold*):

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases} \quad (3.3)$$

ii) Função Linear:

$$\varphi(v) = v \text{ com } 0 < v < \infty \quad (3.4)$$

iii) Função Sigmoidal:

a. Função Logística:

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad (3.5)$$

b. Função Tangente Hiperbólica:

$$\varphi(v) = \tanh(v) \quad (3.6)$$

Na função logística, a é o parâmetro de inclinação da função sigmóide a qual varia no intervalo $[0,1]$, quanto maior o valor do parâmetro, maior a inclinação da função sigmóide. Já a função tangente hiperbólica é para casos onde se admite valores negativos (FAUSETT, 1994, p. 18).

3.2.2.2 Os tipos de topologia das redes neurais

Tão importante para uma Rede Neural quanto sua função de ativação é a arquitetura da rede. São três os principais tipos de arquitetura, cada uma com subvariações que se adaptam a cada tipo de problema. Para Haykin (*apud* BAPTISTELLA, 2005) existem três tipos de arquitetura fundamentalmente diferentes: as redes de camada única alimentadas adiante, as redes de múltiplas camadas alimentadas adiante e as redes recorrentes. Segue uma breve descrição dos três tipos de arquitetura de rede citados:

- i) Redes alimentadas adiante (*feed-forward*) com camada única: a camada de entrada não é considerada, pois não é feito cálculo algum, logo o termo “camada única” refere-se à camada de saída, conforme a figura 3.4;
- ii) Redes alimentadas adiante (*feed-forward*) com múltiplas camadas: a camada de entrada continua não sendo considerada, pois não é feito cálculo algum, portanto somente são contadas as camadas chamadas intermediárias e a camada de saída, conforme a figura 3.5;
- iii) Redes alimentadas adiante com laços de realimentação (redes recorrentes): o que diferencia este tipo de rede dos outros dois tipos de rede é o fato de haver pelo menos um laço de realimentação, conforme a figura 3.6.

FIGURA 3.4 - ARQUITETURA DE CAMADA ÚNICA

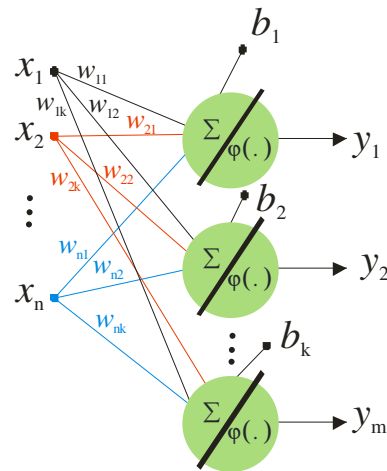


FIGURA 3.5 - ARQUITETURA DE MÚLTIPLAS CAMADAS

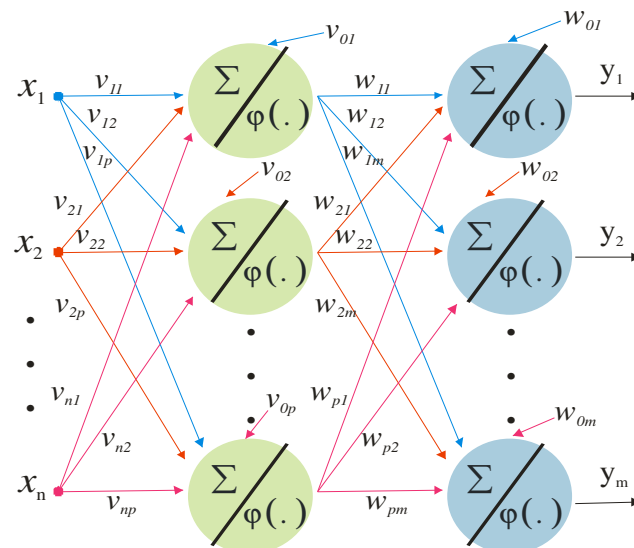
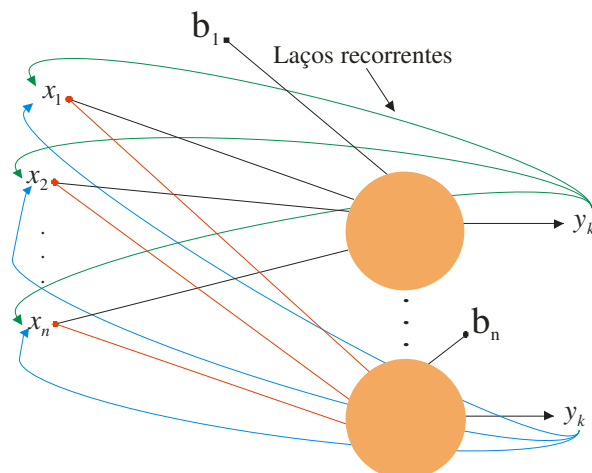


FIGURA 3.6 - ARQUITETURA RECORRENTE



3.2.2.3 O aprendizado das redes neurais artificiais

As Redes Neurais Artificiais “aprendem” por experiência, por meio de dados históricos que contenham os atributos escolhidos para o problema. A partir dos dados históricos, define-se o conjunto de treinamento contendo dois terços dos dados e o conjunto de testes com o restante dos dados, tal procedimento é conhecido como *hold-out*.

Existe também outra versão deste procedimento, na qual o conjunto de treinamento possui metade dos dados, o conjunto de testes um quarto dos dados e um novo conjunto, chamado de conjunto de validação, possui o restante dos dados. O aprendizado geralmente se constitui no ajuste do conjunto de pesos de modo que a rede consiga executar uma tarefa específica, podendo ser realizado, basicamente, de duas formas distintas de acordo com Haykin (2001):

- **Aprendizado supervisionado** - é aquele que utiliza um conjunto de pares (entrada-saída), em que para cada padrão de entrada é especificado um padrão de saída desejado (resposta desejada). O aprendizado ocorre no momento em que a saída gerada pela rede, a partir dos cálculos efetuados com o padrão de entrada e os pesos correntes, for diferente da saída desejada e o algoritmo de treinamento, segundo alguns critérios, ajusta os pesos da rede de forma a reduzir o erro. Essa dinâmica é repetida para todo o conjunto de dados (entradas e saídas) inúmeras vezes, até que a taxa atinja uma faixa considerada satisfatória.
- **Aprendizado não-supervisionado** - este tipo de aprendizado também é conhecido como aprendizado auto-supervisionado, e classifica os padrões similares sem utilizar pares (entrada - saída), isto é, no treinamento da rede são usados apenas valores de entrada. A rede trabalha essas entradas e as organiza de modo a classificá-las mediante algum critério de semelhança. Esse tipo de rede utiliza os neurônios como classificadores, e os dados de entrada como elementos de classificação.

3.2.3 O Algoritmo *Back-propagation*

Um dos algoritmos de treinamento mais utilizados, por ter um desempenho altamente satisfatório em máquinas não tão potentes, é o algoritmo *back-propagation*.

O treinamento de uma rede via *back-propagation* possui três estágios: a apresentação dos padrões à rede, a comparação do valor obtido com o valor desejado, a retro-propagação

do erro e o ajuste dos pesos sinápticos. Após o treinamento, quando a rede já “aprendeu” a classificar os padrões, a aplicação da rede resume-se em apresentar os padrões à rede para serem classificados (primeira fase).

Um detalhe importante a ser mencionado é o de que uma rede de camada única utilizando *back-propagation* é altamente limitada no aprendizado, resolvendo uma quantidade muito pequena de problemas. Já uma rede de múltiplas camadas (uma ou mais camadas escondidas) pode resolver uma quantidade muito maior de problemas satisfatoriamente.

Pode-se utilizar mais de uma camada escondida na arquitetura da rede, entretanto o custo computacional aumenta demasiadamente. Como não há uma perda com o uso de apenas uma camada escondida de acordo com o teorema de Kolmogorov, esta é a prática mais utilizada. A função de ativação de uma Rede Neural que utiliza o algoritmo *back-propagation* é, geralmente, uma daquelas contidas nas equações (3.3), (3.4), (3.5) e (3.6).

Neste trabalho, para a camada escondida, foi utilizada a função logística e, para a camada de saída, a função linear mostrou um melhor desempenho.

3.2.3.1 O algoritmo de treinamento *back-propagation*

No algoritmo descrito em Fausett (1994) a notação utilizada é a seguinte:

x : Vetor de entrada do conjunto de treinamento: $x = (x_1, \dots, x_i, \dots, x_n)$;

t : Vetor com os respectivos valores-alvo de saída: $t = (t_1, \dots, t_i, \dots, t_m)$;

δ_k : Porção do erro a ser ajustado no peso w_{jk} devido ao erro na saída Y_k ;

δ_j : Porção do erro a ser ajustado no peso v_{ij} devido à retro-propagação do erro da camada de saída à camada escondida Z_j ;

α : Taxa de aprendizagem ($0 < \alpha < 1$);

X_i : Entrada para o padrão i ;

Z_j : Neurônio escondido j , a entrada para Z_j é denotada por z_in_j e saída é denotada por $z_j = f(z_in_j)$;

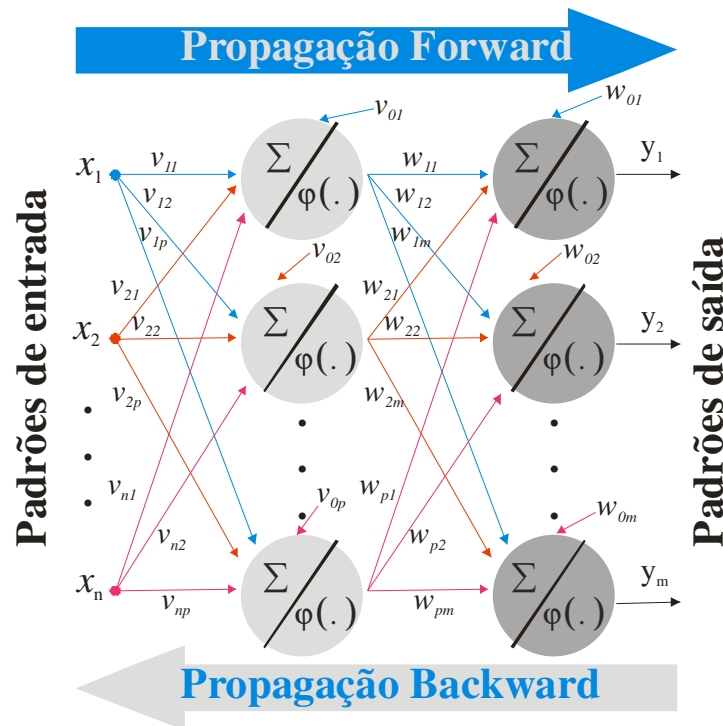
v_{oj} : Vício na saída j ;

w_{0k} : Vício na saída k ;

Y_k : Neurônio de saída k , a entrada para Y_k é denotada por y_in_k e a saída é denotada por $y_k = f(y_in_k)$.

A figura 3.7 tenta ilustrar o funcionamento de uma rede neural *feed-forward* que esteja utilizando o algoritmo *back-propagation*, com apenas uma camada escondida.

FIGURA 3.7 - ESQUEMA DO ALGORITMO *BACK-PROPAGATION*



A validação do método tem a ver com o gradiente da função de ativação do modelo “a base matemática para o algoritmo *back-propagation* é a técnica de otimização conhecida como gradiente decrescente” (FAUSETT, 1994, p. 296). O algoritmo de treinamento *back-propagation* está descrito em Fausett (1994), como segue:

1. Inicialização dos pesos (valores randômicos);
2. Enquanto a condição de parada é falsa, fazer os passos 2.1.-2.2.;
 - 2.1. Para cada par do conjunto de treinamento, fazer os passos 2.1.1.-2.1.5.;

2.1.1. Cada neurônio na camada de entrada ($X_i, i = 1, \dots, n$) recebe o padrão x_i e o transmite para todos os neurônios na camada seguinte (os neurônios da camada escondida);

2.1.2. Cada neurônio na camada escondida ($Z_j, j = 1, \dots, p$) soma seu sinal de peso de entrada, $z_in_j = v_{0j} + \sum_{i=1}^n x_i v_{ij}$, aplica sua função de ativação para calcular seu sinal de saída, $z_j = f(z_in_j)$, e envia esse sinal a todos os neurônios na camada seguinte (unidades de saída);

2.1.3. Cada neurônio na camada de saída ($Y_k, k = 1, \dots, m$) soma seus sinais de pesos de entradas, $y_in_k = w_{0k} + \sum_{j=1}^p z_j w_{jk}$, e aplica sua função de ativação para calcular seu sinal de saída, $y_k = f(y_in_k)$;

2.1.4. Retro-propagação do erro:

2.1.4.1. Cada neurônio na camada de saída ($Y_k, k = 1, \dots, m$) recebe um valor alvo correspondente ao valor de entrada, calcula o valor do erro, $\delta_k = (t_k - y_k) f'(y_in_k)$, calcula o termo de correção do peso (utilizado para atualizar w_{jk} posteriormente), $\Delta w_{jk} = \alpha \delta_k z_j$, calcula o termo de correção “vício” (utilizado para atualizar w_{0k} posteriormente), $\Delta w_{0k} = \alpha \delta_k$, e envia δ_k para os neurônios da camada anterior;

2.1.4.2. Cada neurônio da camada escondida ($Z_j, j = 1, \dots, p$) soma seus valores de delta (dos neurônios da camada de saída), $\delta_in_j = \sum_{k=1}^m \delta_k w_{jk}$, multiplica pela derivada da função de ativação para calcular o termo de correção do erro, $\delta_j = \delta_in_j f'(z_in_j)$, calcula seu termo de correção do peso (utilizado para atualizar v_{ij} posteriormente), $\Delta v_{ij} = \alpha \delta_j x_i$, e calcula seu termo de correção do “vício” (utilizado para atualizar v_{0j} posteriormente), $\Delta v_{0j} = \alpha \delta_j$;

2.1.5. Atualização dos pesos e dos “vícios”:

2.1.5.1. Cada neurônio na camada de saída ($Y_k, k = 1, \dots, m$) atualiza seus pesos e vícios ($j = 0, \dots, p$): $w_{jk}(\text{novo}) = w_{jk}(\text{velho}) + \Delta w_{jk}$;

2.1.5.2. Cada Neurônio da camada escondida ($Z_j, j = 1, \dots, p$) atualiza seus pesos e vícios ($i = 1, \dots, n$): $v_{ij}(\text{novo}) = v_{ij}(\text{velho}) + \Delta v_{ij}$;

2.2. Testar a condição de parada.

3.2.3.2 Parâmetros Utilizados no Algoritmo *back-propagation*

1. A função de ativação para a camada escondida foi a função logística que pode ser vista na equação (3.5) e possui derivada:

$$\varphi'(v) = \frac{ae^{-av}}{(1 + e^{-av})^2} \quad (3.7)$$

Fazendo $y = \varphi(v)$, tem-se $e^{-av} = \frac{1}{y} - 1$ e, conseqüentemente:

$$\varphi'(v) = ay(1 - y) \quad (3.8)$$

Para um neurônio localizado na camada de saída, costuma-se mudar a notação $y(n)$ para $out(n)$ e, assim tem-se o gradiente local do neurônio expresso por:

$$\begin{aligned} \delta &= e\varphi'(v) \\ \delta &= (d - out)a \cdot out(1 - out) \end{aligned} \quad (3.9)$$

Caso o neurônio esteja numa camada escondida, a notação volta a ser $y(n)$ e, assim o gradiente local do neurônio é representado por:

$$\begin{aligned} \delta_j &= \varphi'(v) \sum_k \delta_k \cdot w \\ \delta_j &= ay(1 - y) \sum_k \delta_k \cdot w \end{aligned} \quad (3.10)$$

2. A função de ativação utilizada na camada de saída foi a identidade, com o intervalo modificado de $(-\infty, \infty)$ para $(0, \infty)$ em virtude da não haver preços negativos para imóveis;
3. A taxa de aprendizagem da rede foi de $\gamma = 0,01$, pois com taxas maiores a convergência se tornava difícil;
4. O número de neurônios na camada escondida foi variável, dependendo do desempenho da rede, mas foram efetuados testes de um até o limite de 50 neurônios na camada escondida (em nenhum caso foram precisos os 50 neurônios);
5. O erro máximo permitido na classificação foi de 10^{-5} , escolhido arbitrariamente;
6. O número máximo de iterações ficou definido, arbitrariamente, como sendo 2000 iterações.

3.2.3.3 Variações no algoritmo *back-propagation*

Para acelerar a convergência da rede, ou seja, acelerar a diminuição do erro e a minimização do gradiente, normalmente são feitas algumas alterações no funcionamento do algoritmo *Back-propagation*. Neste trabalho foram feitas duas variações:

1. Taxa de Momento: A taxa de momento (*momentum*) μ serve para direcionar a atualização dos pesos da rede de modo que esta seja uma combinação do gradiente anterior e do gradiente da iteração atual, esta taxa deve estar no intervalo $[0,1]$. De acordo com Fausett (1994) a convergência as vezes é mais rápida quando a taxa de momento é adicionada às fórmulas de atualização dos pesos. De uma forma simples, é possível representar um peso no passo $(t+1)$ dependendo dos pesos dos passos t e $(t-1)$ na equação (3.11):

$$w(t+1) = w(t) + ay\delta + \mu[w(t) - w(t-1)] \quad (3.11)$$

2. Outra variação adotada neste trabalho foi a taxa de aprendizagem adaptativa: se a *performance* da rede diminui (em relação à iteração anterior), a taxa de aprendizagem aumenta sendo multiplicada por um fator $inc = 1,05$, chamado de fator de incremento; se a performance da rede aumenta de maneira exagerada (104% ou mais em relação à iteração anterior, padrão do MATLAB, podendo ocasionar *overfitting*), a taxa de

aprendizagem diminui via um fator de decremento $dec = 0,7$. Todos os valores aplicados nesta variação do algoritmo *Back-propagation* são padrões do MATLAB. Vale observar que diversos critérios heurísticos são propostos na literatura, como, por exemplo, a regra *delta-bar-delta* (FAUSETT, 1994), onde a principal característica é que cada peso da rede tem sua própria taxa de aprendizagem, já Haykin (2001), descreve outros 4 critérios heurísticos, sendo o primeiro o próprio *delta-bar-delta*.

3.2.4 Redes Competitivas

A principal característica das redes competitivas é o seu tipo de aprendizado, o qual é do tipo não-supervisionado. Ao contrário das redes baseadas na retro-propagação do erro, as soluções encontradas normalmente não são as ótimas, bastando apenas que apresentem resultados satisfatórios e seja factível.

A aprendizagem não-supervisionada modifica repetidamente os pesos sinápticos até que haja uma configuração final que seja favorável aos critérios estabelecidos inicialmente, sendo fornecido a elas apenas padrões. Tal fato é possível, pois já foi demonstrado por Turing (*apud* HAYKIN, 2001) “uma ordem global pode surgir de interações locais”, o que quer dizer que resultados globais podem ser encontrados localmente.

Existem quatro princípios básicos da auto-organização Arruda (2004 *apud* FRANCISCO, 2004):

- Princípio 1: a modificação dos pesos tende a se auto-amplificar: baseada nos sinais locais (pré e pós) sinápticos, os pesos são reforçados caso os sinais obtidos antes e depois da sinapse sejam coincidentes e também, caso o padrão de resposta seja coincidente ao de entrada. Desta forma, todos os pesos associados são reforçados e a estabilidade é alcançada via competição: se alguns pesos aumentam, outros tendem a desaparecer.
- Princípio 2: princípio da plasticidade das sinapses: a limitação de recursos leva a competição entre sinapses e, assim, o problema de seleção das sinapses mais forte melhora às custas das outras.
- Princípio 3: a modificação dos pesos tende a cooperar: as vizinhanças dos neurônios mais fortes são melhoradas, apesar do aspecto global de competição da rede, todos os neurônios são igualmente estimulados.

- Princípio 4: princípio da redundância de informação: a ordem e a estrutura da rede devem determinar a redundância de informação necessária para garantir a aquisição de conhecimento; se existirem poucos neurônios, a rede pode aprender e desaprender, é o princípio de que quanto maior o espaço, mais coisas pode-se guardar.

Existem dois tipos de redes competitivas:

1. Peso fixo: não ocorre o treinamento, somente a competição:
 - a. MAXNET (ou “*WINNER TAKES ALL*”) – utilizada para tomar decisões e encontrar o maior valor, normalmente é acoplada em outra rede para determinação do neurônio vencedor, onde o vencedor leva tudo.
 - b. Chapéu Mexicano – os neurônios possuem pesos positivos e negativos.
 - c. Rede de *Hamming* – é um classificador de máxima verossimilhança, pode ser classificada também como uma rede probabilística e sua técnica de agrupamento utiliza a distância de *Hamming*.
2. Peso variável: ocorre a competição e o treinamento, onde é utilizado o princípio de aprendizagem de Kohonen.
 - a. Mapa de Kohonen – algoritmo de agrupamento que será detalhado a seguir.
 - b. Redes com quantização do vetor de aprendizagem (LVQ – *Learning Vector Quantization*) – utilizada para classificação de padrões, sendo este tipo de rede com aprendizado supervisionado.
 - c. Redes com contra-propagação – semelhante as redes de múltiplas camadas com uma estrutura competitiva.

3.2.4.1 Redes de Kohonen

As Redes de Kohonen são do tipo SOM – *Self-Organizing Maps* (Mapas Auto-Organizáveis); são redes baseadas em competição. A principal diferença entre as redes baseadas em competição e as redes que utilizam o algoritmo *Back-propagation* é que o aprendizado daquelas é não-supervisionado, entretanto, seu funcionamento é bem mais simples do que as redes com *back-propagation*.

O objetivo principal da Rede de Kohonen é idêntico ao da técnica estatística conhecida como *Clustering Analysis* (Análise de Agrupamento), técnica utilizada nos trabalhos de Alves (2005) e Braulio (2005).

De acordo com Francisco (2004) o processo de aprendizagem é baseado no aprendizado competitivo e as modificações dos pesos sinápticos são confinadas à vizinhança do neurônio ativado. A ordem global, ou seja, o equilíbrio da rede, surge da interação entre neurônios vizinhos. Os neurônios de saída competem entre si para serem ativados, de forma que apenas um neurônio de saída seja considerado “vencedor”.

Os neurônios em uma Rede de Kohonen são colocados nos nós de uma grade de maneira a assumir uma forma geométrica conhecida (retângulo, hexágono, etc.), que é usualmente de uma ou duas dimensões, em virtude de mapas tridimensionais possuírem difícil aplicação e compreensão. Cada neurônio na grade é completamente conectado a todos os neurônios da camada de entrada.

Os padrões de entrada são apresentados à rede e, a cada padrão apresentado, tem-se uma região de atividade na grade. A localização e natureza de uma determinada região variam de um padrão de entrada para outro.

Dessa forma, todos os neurônios da rede devem ser expostos a um número suficiente de diferentes padrões de entrada, garantindo assim que o processo de auto-organização ocorra de forma apropriada. Os neurônios recebem padrões de entrada durante o processo competitivo de aprendizado. Serão considerados “vencedores” os neurônios que mais se assemelharem ao padrão de entrada, sendo que para esta comparação são utilizadas medidas de distâncias, normalmente a distância euclidiana.

Outra forma de avaliar a semelhança entre o padrão de entrada e o neurônio vencedor é o produto interno dos vetores de entrada e peso normalizados, que pode ser interpretado como a correlação entre o vetor de entrada e o vetor peso (FAUSETT, 1994).

Para Francisco (2004) a localização destes neurônios vencedores se torna ordenada de tal forma que um sistema de coordenadas é criado na grade, para diferentes características de entrada. Assim sendo, a rede é caracterizada pela formação de um mapa topográfico dos padrões de entrada, no qual as localizações espaciais (ou coordenadas) dos neurônios na grade são indicativas de características dos padrões de entrada.

Neste trabalho, Rede de Kohonen possui apenas duas camadas: a camada de entrada e a de saída. A camada de entrada é conectada a um vetor de entrada e camada de saída possui

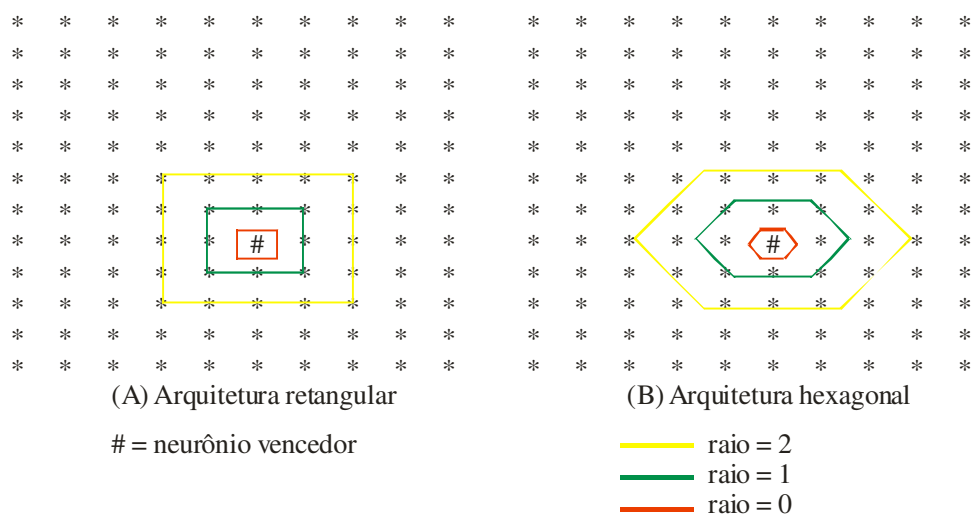
vários neurônios formando uma grade retangular. Antes da aplicação propriamente dita deste tipo de rede é necessário determinar:

1. a taxa de aprendizagem e o tamanho raio topológico (para definir quantos neurônios serão atualizados); vale lembrar que ambos decrescem conforme o treinamento é executado;
2. a topologia da camada de saída, ou seja, a forma com a qual ela se parecerá. Desta forma fica definido quais neurônios pertencerão à vizinhança do neurônio vencedor.

A figura 3.8 apresenta dois exemplos de topologia para a camada de saída, uma em formato retangular (A) e outra em formato hexagonal (B), também estão definidos os raios topológicos e o neurônio vencedor. Vale salientar que apenas os neurônios alcançados pelo raio topológico têm seus pesos sinápticos atualizados após a iteração.

O algoritmo responsável pela formação do mapa de Kohonen primeiramente inicializa os pesos sinápticos da rede (geralmente aleatórios), posteriormente é definido o raio topológico inicial e, por último, a taxa de aprendizagem e, só então, o processo de formação do Mapa de Kohonen acontece.

FIGURA 3.8 - PARÂMETROS TOPOLÓGICOS DE VIZINHANÇA E RAIOS TOPOLÓGICOS



Para a formação do Mapa de Kohonen são consideradas três etapas detalhadas a seguir, com base em Francisco (2004):

1. Competição: para cada padrão de entrada, os neurônios calculam um determinado valor com o intuito de discriminar os padrões. Em geral, o neurônio com o maior valor calculado é o neurônio vencedor; no caso das distâncias é o neurônio com menor valor. Ainda, é possível utilizar o inverso da distância, então o maior valor é o vencedor;
2. Cooperação: o neurônio vencedor define, por meio do raio topológico, quais neurônios terão o “privilégio” de terem seus pesos sinápticos atualizados, promovendo assim a cooperação entre neurônios vizinhos;
3. Adaptação sináptica: este dispositivo permite que os neurônios do raio topológico (neurônios excitados) ajustem seus pesos de maneira a aumentar seu valor calculado na primeira fase na iteração subsequente.

3.2.4.1.1 Tipos de distância utilizados

Outra característica importante a ser determinada num mapa auto-organizável de Kohonen é o tipo de distância a ser utilizado. Obviamente, existem inúmeros tipos de distância, porém, as três mais utilizadas são:

1. a distância euclidiana entre dois vetores x e y , é dada na equação (3.12);

$$d(x, y) = \sqrt{\sum (x - y)^2} \quad (3.12)$$

2. a distância de ligação, de um vetor x a um vetor y , de um conjunto de vetores S , está descrita na equação (3.13);

$$d(x_i, y_j) = \begin{cases} 0, & \text{se } i = j; \\ 1, & \text{se } \sqrt{\sum (x - y)^2} \leq 1; \\ 2, & \text{se existe } k, d(x_i, y_k) = d(x_k, y_j) = 1; \\ 3, & \text{se existe } k1, k2, d(x_i, y_{k1}) = d(x_{k1}, y_{k2}) = d(x_{k2}, y_j) = 1; \\ n, & \text{se existe } k1, \dots, kn, d(x_i, y_{k1}) = d(x_{k1}, y_{k2}) = \dots = d(x_{kn}, y_j) = 1; \\ S, & \text{se nenhuma das condições é satisfeita.} \end{cases} \quad (3.13)$$

3. e a distância de *Manhatan* (ou *city-block*), entre dois vetores x e y , está na equação (3.14).

$$d(x, y) = \sum |x - y| \quad (3.14)$$

3.3 MÉTODOS ESTATÍSTICOS MULTIVARIADOS

3.3.1 Introdução

Esta seção tem o objetivo de destacar, bem como, expor os principais conceitos dos Métodos Estatísticos Multivariados utilizados como auxílio à aplicação das técnicas de Redes Neurais desenvolvidas neste trabalho.

Quando se trata de Estatística Univariada (apenas uma variável), normalmente o vetor de observações é representado da forma:

$$X_m = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad (3.15)$$

Na equação matricial (3.15), m é o número de observações para que se possa calcular qualquer estatística com a amostra X_m , seja Média, Desvio-padrão, entre outras.

A mudança para Estatística Multivariada é apenas uma expansão, pura e simples, do conceito apresentado na equação matricial (3.15), pois a forma de se obter as estatísticas são extremamente similares. Normalmente, a matriz de observações é representada conforme a equação (3.16), com as variáveis dispostas em colunas e as observações em linhas. Entretanto, alguns autores preferem inverter esta disposição, mas isto não gera qualquer tipo de problema conceitual.

$$X_{m \times p} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mp} \end{bmatrix} \quad (3.16)$$

A matriz da equação (3.16) possui p vetores (vetores aleatórios), cada vetor representa uma variável aleatória X , ou seja, a matriz possui p variáveis aleatórias, cada uma com m observações. Cada vetor aleatório pode ser representado individualmente da forma:

$\underline{X}_1 = [X_{11}, X_{21}, \dots, X_{m1}]^T = [x_{11}, x_{21}, \dots, x_{m1}]^T$, para representar a primeira variável (primeira coluna), $\underline{X}_2 = [X_{12}, X_{22}, \dots, X_{m2}]^T = [x_{12}, x_{22}, \dots, x_{m2}]^T$, para a segunda variável (segunda coluna), até a representação da última variável $\underline{X}_p = [X_{1p}, X_{2p}, \dots, X_{mp}]^T = [x_{1p}, x_{2p}, \dots, x_{mp}]^T$ que estaria disposta na última coluna.

3.3.2 Estatísticas Descritivas Multivariadas

As informações amostrais podem ser resumidas em números sumários conhecidos como estatísticas e que podem constar nas observações multivariadas $[\underline{x}_1, \underline{x}_2, \underline{x}_3, \dots, \underline{x}_n]$.

As estatísticas são usadas na inferência sobre os parâmetros, ou seja, na estimação do vetor médio $\underline{\mu}$, da matriz de covariância $\underline{\Sigma}$ ou da matriz de correlação $\underline{\rho}$, entre outros. Desta forma o vetor médio populacional $\underline{\mu}$ deve ser estimado pelo vetor médio amostral, $\underline{\bar{X}}$, definido pela expressão contida na equação (3.17).

$$\underline{\bar{X}} = \frac{1}{m} \cdot \sum_{i=1}^m \underline{X}_i \quad (3.17)$$

Na equação (3.17) acima, \underline{X}_i com $i = 1, 2, \dots, m$, corresponde às observações amostrais do vetor \underline{X} e m é o tamanho da amostra observada. Outros parâmetros de uma população multivariada $f(\underline{x})$ podem ser avaliados, tais como a matriz de covariância $\underline{\Sigma}$, definida na equação (3.18).

$$\underline{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix} \quad (3.18)$$

Na equação (3.18), na diagonal principal, estão dispostas as variâncias das variáveis aleatórias e, fora da diagonal principal, as covariâncias entre elas. A matriz de correlação $\underline{\rho}$, está definida na equação (3.19).

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \quad (3.19)$$

As correlações entre as variáveis estão fora da diagonal principal. Então, estes parâmetros, \sum e ρ , são estimados, respectivamente, pela matriz de covariância amostral S e pela matriz de correlação amostral R , cujas expressões estão representadas nas equações (3.20) e (3.21).

$$S = \frac{1}{m-1} \cdot \sum_{i=1}^m (\underline{X}_i - \overline{\underline{X}})(\underline{X}_i - \overline{\underline{X}})^T = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{bmatrix} \quad (3.20)$$

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (3.21)$$

Na equação (3.22), está representada variância amostral da j -ésima variável aleatória \underline{X}_j , na equação (3.23), a covariância amostral entre as variáveis \underline{X}_j e \underline{X}_k .

$$s_j^2 = \frac{1}{m-1} \cdot \sum_{i=1}^m (X_{ij} - \overline{X}_j)^2 \quad (3.22)$$

$$s_{jk} = \frac{1}{m-1} \cdot \sum_{i=1}^m (X_{ij} - \overline{X}_j)(X_{ik} - \overline{X}_k) \quad (3.23)$$

$$r_{jk} = \frac{s_{jk}}{s_j \cdot s_k} \quad (3.24)$$

A equação (3.24) mostra como são calculadas as correlações fora da diagonal principal, ou seja, o quociente entre a covariância amostral e o produto dos desvios-padrões amostrais para $j \neq k$. Esses estimadores são os melhores para se determinar os parâmetros.

Os primeiros são EUMV (Estimador Uniformemente de Mínima Variância) e o último é EMV (Estimador de Máxima Verossimilhança).

3.3.2.1 Padronização de dados multivariada

Um fato que ocorre em muitos casos, inclusive na avaliação de imóveis, é o de que as variáveis da matriz de dados X podem estar dispostas em escalas diferentes, podendo ocasionar erros de arredondamento, em virtude de a matriz estar “mal condicionada”.

Para evitar este problema, uma prática quase obrigatória é a padronização dos dados, de acordo com a equação (3.25), onde \bar{X}_j e S_j são, respectivamente, a média e desvio-padrão da j -ésima variável e X_{ij} é a i -ésima observação da j -ésima variável.

$$Z_j = \frac{X_{ij} - \bar{X}_j}{S_j} \quad (3.25)$$

3.3.3 Análise de Componentes Principais

O método da Análise de Componentes Principais procura explicar (representar) a estrutura da variância e covariância de um vetor aleatório através de poucas combinações lineares das variáveis originais.

Seu objetivo principal é evitar o problema clássico da multicolinearidade dos dados, pois consiste numa transformação, de eixos, tornando as novas variáveis (combinações lineares) não correlacionadas. Assim, uma matriz de dados X de ordem $n \times p$ pode ser substituída por outra de ordem $n \times m$ sendo $m \leq p$.

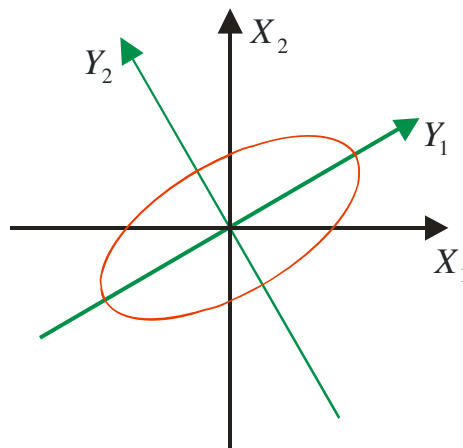
Embora as p componentes sejam necessárias para reproduzir toda a variabilidade presente na estrutura de covariância do vetor \underline{X} de dimensão p , freqüentemente, uma grande parte desta variabilidade poderá ser explicada por um número $m < p$ de Componentes Principais.

Neste caso, existe praticamente a mesma quantidade de informações nas m Componentes Principais do que nas p variáveis originais. A Análise de Componentes Principais freqüentemente revela relações que não eram previamente consideradas e assim permitem interpretações que não iriam, de outro modo, aparecer.

3.3.3.1 Componentes principais populacionais

Algebricamente, as componentes principais são combinações lineares das p variáveis originais $[X_1, X_2, \dots, X_p]$ que compõem o vetor aleatório \underline{X} . Geometricamente, as combinações lineares representam a seleção de um novo sistema de coordenadas, obtido por rotação do sistema original, sendo que os novos eixos representam as direções com variabilidade máxima. Como exemplo, tem-se, na figura 3.9, a representação da estrutura de componentes principais para $p = 2$.

FIGURA 3.9 - ANÁLISE DE COMPONENTES PRINCIPAIS



Na figura 3.9, X_1 e X_2 são os eixos originais, Y_1 e Y_2 são os novos eixos (eixos originais após a aplicação da Técnica da Análise das Componente principais, eixos originais rotacionados, centrados na média amostral).

Obtém-se as Componentes Principais a partir da matriz de covariância ou da matriz de correlação, que resumem a estrutura de relacionamento das p variáveis originais que compõem o vetor \underline{X} .

Então, da matriz de covariância Σ ou da matriz de correlação ρ , obtém-se os autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ e os respectivos autovetores $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$. Com estas informações são construídas as equações que definem as componentes principais, conforme descrito na equação (3.25), na qual $i = 1, 2, \dots, p$.

$$Y_i = \underline{e}_i \cdot \underline{X} \quad (3.26)$$

Importante lembrar que, por definição, as Componentes Principais são combinações lineares não correlacionadas, uma vez que a matriz dos autovetores P , na equação (3.26), é ortogonal, ou seja, $PP' = P'P = I$.

$$P = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix} \quad (3.27)$$

A equação (3.28) especifica o procedimento para se calcular a variância da i -ésima componente principal, lembrando que a covariância entre as componentes principais j e k ($j \neq k$) é nula, ou seja, $\text{cov}(Y_j, Y_k) = 0$.

$$V(Y_i) = V(\underline{e}_i' \underline{X}) = \underline{e}_i' V(\underline{X}) \underline{e}_i = \underline{e}_i' \cdot \sum \cdot \underline{e}_i \quad (3.28)$$

Portanto define-se:

- A primeira componente principal como a combinação linear $Y_1 = \underline{e}_1' \underline{X}$ que maximiza a variância de Y_1 , sob a restrição $\underline{e}_1' \underline{e}_1 = 1$.
- A segunda componente principal como a combinação linear $Y_2 = \underline{e}_2' \underline{X}$ que maximiza $V(\underline{e}_2' \underline{X})$ sujeita à restrição $\underline{e}_2' \underline{e}_2 = 1$ e $\text{cov}(Y_1, Y_2) = 0$.
- A i -ésima componente principal como a combinação linear $Y_i = \underline{e}_i' \underline{X}$ que maximiza $V(\underline{e}_i' \underline{X})$ sujeita à restrição $\underline{e}_i' \underline{e}_i = 1$ e $\text{cov}(Y_i, Y_k) = 0 \forall k \neq i$.

3.3.3.2 Componentes Principais Amostrais

Comumente a matriz de covariância (ou correlação), \sum e/ou ρ , são desconhecidos. Então, a obtenção das componentes principais é feita a partir de seus estimadores, que são a matriz de covariância amostral S , definida na equação (3.28) ou a matriz de correlação amostral R , definida na equação (3.29).

A matriz R depende da matriz D , matriz diagonal dos desvios-padrões amostrais, que está definida na equação (3.30) e o vetor $\underline{\bar{X}}$, vetor de médias das observações amostrais, está definido na equação (3.31).

$$S = \frac{1}{m-1} \cdot \sum (\underline{X}_i - \underline{\bar{X}})(\underline{X}_i - \underline{\bar{X}}) \quad (3.29)$$

$$R = D^{-1}SD^{-1} \quad (3.30)$$

$$D = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p \end{bmatrix} \quad (3.31)$$

$$\underline{\bar{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_3 \end{bmatrix} \quad (3.32)$$

A partir das estimativas dos elementos da estrutura de covariância do vetor aleatório \underline{X} , ou seja, os autovalores $\hat{\lambda}_i$ e os correspondentes autovetores \hat{e}_i e constroem-se as componentes principais amostrais descritas na equação (3.32), equação na qual $i = 1, 2, \dots, p$.

$$\hat{Y}_i = \hat{e}_i \cdot \underline{X} \quad (3.33)$$

3.3.3.3 Critérios para definição do número de componentes principais

Um critério, dos que existem, para a determinação do número de componentes a serem extraídas foi sugerido por Kaiser em 1960. Segundo Johnson e Wichern (2002), Kaiser propôs escolherem-se somente as componentes correspondentes aos autovalores (raízes latentes) de magnitudes maiores do que um.

Outra maneira de se definir o número de componentes é através da percentagem de variação explicada. O pesquisador, neste caso, deve julgar se m componentes explicam suficientemente o relacionamento entre as p variáveis originais. Geralmente, um bom grau de

explicação é superior a 75% para um m pequeno. Em alguns casos, aproveitando-se 98% da variabilidade já ocorre uma redução significativa no número de variáveis.

Conforme mencionado anteriormente, uma propriedade muito importante, talvez a mais importante, das Componentes Principais é a independência entre elas. Desta forma, podem substituir as variáveis originais e eliminar o problema de multicolinearidade.

4 METODOLOGIA PARA A RESOLUÇÃO DO PROBLEMA

Para o desenvolvimento do trabalho e obtenção dos resultados foi desenvolvido um sistema em MATLAB 7.1® (sem recursos gráficos), nominado AIRNA – Avaliação de Imóveis via Redes Neurais Artificiais. A rotina AIRNA é constituída das técnicas: padronização dos dados, mapas auto-organizáveis de Kohonen, Análise de Componentes Principais e Redes *feed-forward* de Múltiplas Camadas (com a utilização do algoritmo *back-propagation*).

Para acelerar a convergência no treinamento, optou-se por utilizar a taxa de aprendizagem adaptativa e taxa de momento (padrões do MATLAB 7.1 ®) e o desempenho do algoritmo melhorou muito, em relação aos testes iniciais, com estas escolhas. A taxa de aprendizagem inicial foi $\gamma=0,01$, sendo a taxa de incremento $inc=1,05$, a taxa de decremento $dec=0,7$ e a taxa de momento $\mu=0,9$. A taxa de incremento é multiplicada pela taxa de aprendizagem, quando o desempenho da rede diminui de uma iteração para outra. O mesmo acontece com a taxa de decremento, quando o desempenho da rede aumenta demasiadamente, podendo ocasionar sobreajuste, conforme descrito no segundo capítulo, anteriormente.

4.1 DESCRIÇÃO DO SISTEMA AIRNA

O sistema AIRNA é composto das seguintes etapas:

1. Leitura e transformação dos dados disponíveis (características do imóveis) contidas em planilhas do *Microsoft Excel*® por meio do comando *xlsread*, comando configurável que lê planilhas e transforma as informações em matrizes no formato do MATLAB. Esta base de dados contém as variáveis dispostas em colunas e as observações em linhas e, inicialmente, não contém as variáveis resposta (que serão lidas posteriormente em outra planilha com o mesmo comando);

2. Após a leitura, os dados são padronizados pelo sistema, pela fórmula $Z_j = \frac{X_{ij} - \bar{X}_j}{S_j}$,

para que haja um melhor desempenho da Rede de Kohonen, pois fica garantido que as variáveis estarão na mesma escala;

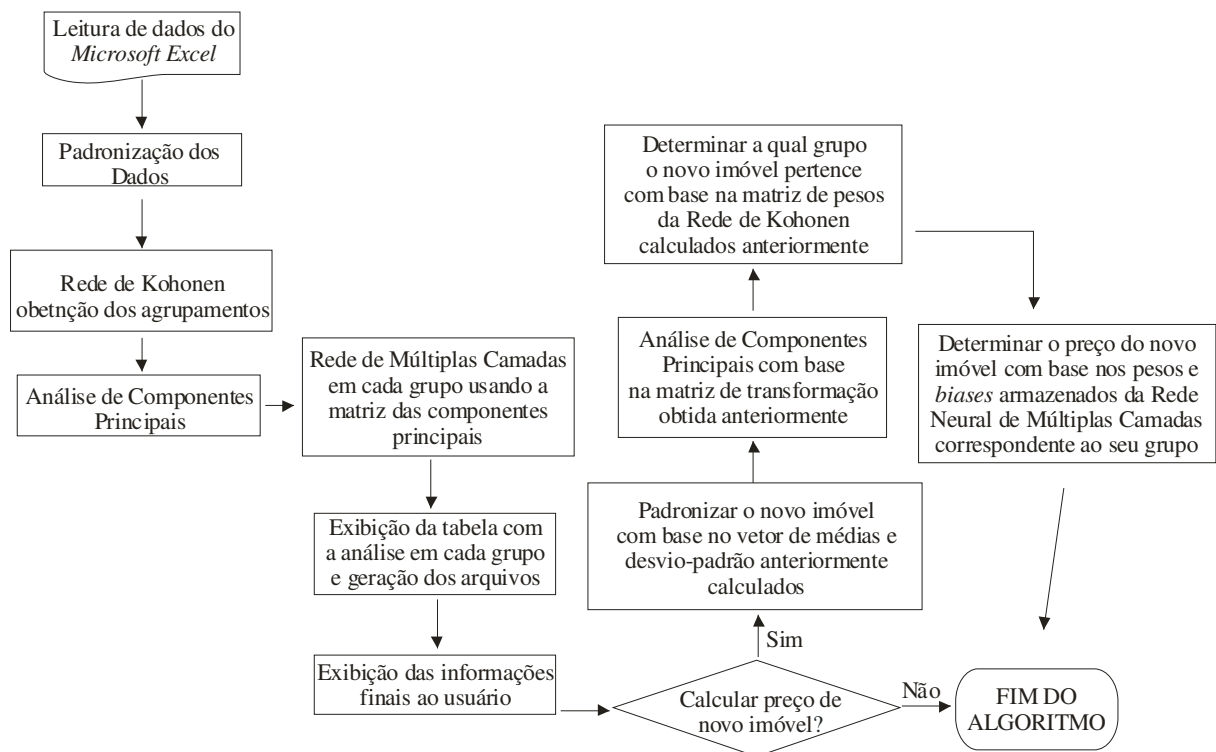
3. Após a padronização, o algoritmo executa a Rede de Kohonen com 300 iterações e com arquitetura escolhida pelo usuário. O sistema recomenda que a arquitetura utilizada seja de 2x2 neurônios (4 neurônios) para uma melhor avaliação da generalização da Rede Neural de Múltiplas Camadas que será executada posteriormente, entretanto outras arquiteturas, com mais neurônios, têm um desempenho ainda melhor no conjunto de treinamento;
 - a. O usuário pode escolher entre dois tipos de arquitetura: retangular (*gridtop*) ou hexagonal (*hextop*);
 - b. O usuário ainda pode escolher entre três tipos de distância: euclidiana (*dist*), distância de ligação (*linkdist*) ou distância de *manhatan* (*mandist*);
4. Após a execução da Rede de Kohonen a matriz de pesos é armazenada.
5. Após a formação dos agrupamentos, o usuário tem a opção de executar (ou não) a Análise de Componentes Principais bem como o de escolher o nível de significância. É extremamente recomendável executar a Análise de Componentes Principais para evitar multicolinearidade;
6. Para cada um dos grupos formados pela Rede de Kohonen, é aplicada uma Rede Neural Multicamadas, com 2000 iterações, variando o número de neurônios na camada escondida até que o erro quadrático máximo permitido seja alcançado ou o número de neurônios na camada escondida seja 50 (que já é uma quantidade não usual para problemas deste porte). As funções de ativação utilizadas são: função logística para a camada escondida e linear, com o intervalo modificado para $[0, \infty)$, para a camada de saída em virtude de esta combinação ter mostrado um desempenho melhor nos testes. Os dados de cada grupo são separados em conjuntos de treinamento e testes, sendo $\frac{3}{4}$ da amostra para o primeiro e $\frac{1}{4}$ da amostra para o segundo conjunto; em virtude do tamanho da amostra ser pequena optou-se por não adotar conjunto de validação;
 - a. Após a execução da Rede Neural Multicamada, são armazenados os pesos finais, bem como os vícios (*biases*) da rede, para cálculos posteriores.
7. Calcula-se a diferença entre o preço de mercado dos imóveis e o preço obtido pela rede neural e uma porcentagem de acerto nos preços com uma tolerância de 10% de

erro é exibida. Em seguida, o usuário é solicitado a pressionar *enter* para executar a rede do próximo grupo até que não existam mais grupos;

8. Após executar a rede em cada grupo, o sistema ainda possibilita o cálculo do preço de um novo imóvel, as características deste novo imóvel são “buscadas” numa planilha previamente salva no diretório especificado no sistema. Também é possível executar separadamente esta última parte do programa, caso o usuário necessite calcular preços de mais de um imóvel;
 - a. O novo imóvel é padronizado com base na média e desvio-padrão dos imóveis avaliados durante a execução do sistema, agrupado de acordo com a matriz armazenada durante a execução da Rede de Kohonen e tem seu preço determinado pelos pesos e vícios (*biases*) obtidos durante a execução do algoritmo *back-propagation*.
9. Após o cálculo do preço do novo imóvel o algoritmo é finalizado.

A seqüência de passos desta rotina está descrita na figura 4.1, um fluxograma explicativo do algoritmo.

FIGURA 4.1 - FLUXOGRAMA EXPLICATIVO DO SISTEMA AIRNA



5 OBTENÇÃO E ANÁLISE DOS RESULTADOS

5.1 OBTENÇÃO DOS AGRUPAMENTOS

Antes de utilizar a técnica de agrupamento, a padronização nos 144 imóveis foi feita separadamente para apartamentos (44), casas (51) e terrenos (24), sendo utilizada a padronização de dados multivariada descrita na seção 3.3.2.1 deste trabalho.

Conforme salientado anteriormente, neste trabalho fez-se o uso da Rede de Kohonen, com 300 iterações para todos os testes, para obtenção dos grupos para cada tipo de imóvel: os 44 apartamentos, as 51 casas e os 24 terrenos. Em cada um dos tipos de imóvel foram obtidos quatro grupos, visto que a quantidade de neurônios utilizada na arquitetura dos mapas auto-organizáveis de Kohonen em todos os casos foi justamente esta. Em nenhum caso o algoritmo gerou menos de quatro grupos, embora isso possa acontecer em outros casos, pois existem técnicas para redução da quantidade de grupos no caso mapas auto-organizáveis maiores.

O sistema desenvolvido possibilita seis combinações no que diz respeito à arquitetura da Rede de Kohonen e a função distância utilizada. Os “pares” são: {(retangular, distância euclidiana), (retangular, distância de ligação), (retangular, distância de Manhattan), (hexagonal, distância euclidiana), (hexagonal, distância de ligação), (hexagonal, distância de Manhattan)}. A comparação direta entre os resultados deste trabalho e os resultados encontrados em Alves (2005) e Braulio (2005) é feita para o melhor resultado dentre as seis combinações e também com os melhores resultados nos trabalhos citados.

A configuração que mostrou melhor resultado (que será apresentado na seção 5.3) para os apartamentos analisados foi aquela composta de arquitetura retangular e a função de distância euclidiana. Esta escolha, conforme mencionado anteriormente, resultou em quatro grupos para os 44 apartamentos avaliados. No primeiro grupo concentraram-se 15 apartamentos, no segundo, 10 apartamentos, no terceiro, três apartamentos e no quarto, 16 apartamentos.

Para as casas, grupo com a maior amostra do problema, a configuração de melhor resultado (seção 5.3) foi a composta de arquitetura hexagonal e a função de distância *manhatan* (ou *city-block*). Assim como no caso dos apartamentos foram obtidos quatro grupos para as 51 casas avaliadas. O primeiro grupo ficou com 13 casas, o segundo também com 13 casas, o terceiro com 11 casas e o quarto, 14 casas.

Para os terrenos, grupo com a menor amostra do trabalho, a configuração mais satisfatória (seção 5.3) foi a composta de arquitetura retangular e a função de distância *manhatan* (ou *city-block*). Não diferindo dos tipos de imóveis anteriores, também foram obtidos quatro grupos de terrenos para os 24 terrenos avaliados. O primeiro grupo contém seis terrenos, o segundo grupo cinco terrenos, o terceiro, concentrou cinco terrenos e o quarto ficou com oito terrenos apenas.

5.2 COMPONENTES PRINCIPAIS

A técnica da Análise de Componentes Principais foi aplicada, principalmente para evitar o problema clássico da multicolinearidade e, também, para tentar diminuir a quantidade de variáveis.

O nível de explicação da variabilidade dos dados para todos os casos foi 98% da variabilidade total dos dados, ou seja, apenas 2% de “perda de informação”. As tabelas 5.1 a 5.3 mostram os escores resultantes da Análise de Componentes Principais obtidos para os tipos de imóveis urbanos avaliados. Nas tabelas as colunas correspondentes a cada componente principal estão enumeradas e a coluna dos grupos pela abreviação G.

No caso dos apartamentos, que inicialmente possuíam 21 características, os 98% da variabilidade dos dados explicados puderam ser condensados em apenas 11 componentes principais. Já para as casas, que inicialmente possuíam 18 variáveis, foram obtidas 12 componentes principais com a mesma significância dos apartamentos. Para os terrenos, que possuíam 10 variáveis, com a mesma significância dos tipos de imóveis anteriores, foram obtidas oito componentes principais.

TABELA 5.1 - COMPONENTES PRINCIPAIS DOS 44 APARTAMENTOS

continua											G
1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	6 ^a	7 ^a	8 ^a	9 ^a	10 ^a	11 ^a	
-4,55	-0,41	0,03	1,07	-0,88	1,20	0,51	0,23	0,57	-0,21	0,13	1
-4,76	-0,62	-1,55	0,78	-0,84	0,37	-0,35	0,20	0,52	-1,41	0,38	1
-3,95	-0,15	0,62	-0,19	-0,38	1,28	0,46	-0,52	0,24	-0,80	0,33	1
-3,74	5,58	-2,38	-2,75	0,28	-2,66	2,45	-1,62	1,07	0,66	-0,76	1
-2,60	-1,64	-1,95	-0,50	-0,09	-0,68	-0,23	1,67	-0,23	0,23	-0,62	1
-1,53	-0,27	-3,01	0,65	1,33	-0,37	-0,06	-0,17	0,19	-0,73	0,81	1
-2,13	-0,66	-0,16	1,36	1,38	-0,48	-0,93	-1,33	-0,03	-1,33	0,78	1
-1,51	0,82	-2,10	1,07	1,10	0,09	-0,27	0,82	-0,53	1,71	1,42	1
-5,02	1,45	1,47	0,29	-1,22	0,36	-0,49	0,39	0,02	-0,06	0,15	1

conclusão											G
1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	9ª	10ª	11ª	
-1,56	-1,20	-1,47	0,63	0,32	1,02	1,02	-0,25	-2,01	0,41	-0,57	1
-1,19	-1,70	-1,15	1,75	1,71	-1,86	-0,86	0,09	-0,19	-0,28	-0,63	1
-4,83	2,57	1,83	0,09	-0,87	0,56	-0,78	0,72	-0,87	-0,27	-0,17	1
-3,94	2,53	2,10	-1,40	-0,42	0,67	-1,02	0,33	-0,40	-1,04	-0,32	1
-1,14	-1,77	-0,61	0,02	1,60	-0,01	0,36	0,42	-0,22	0,62	-0,53	1
-0,97	-1,77	-0,32	-0,35	1,77	0,62	0,46	-0,22	0,00	0,19	-0,39	1
-0,99	-0,78	0,92	0,52	1,26	1,08	0,59	-0,30	0,87	0,77	-0,09	2
-0,82	-0,78	1,21	0,16	1,43	1,71	0,69	-0,93	1,09	0,35	0,05	2
-0,25	-0,62	1,01	-1,21	1,17	-0,21	-0,24	-0,11	-0,22	0,54	0,11	2
-0,65	-0,32	1,42	0,06	0,82	0,34	0,05	-0,42	-0,52	0,99	0,39	2
1,29	0,37	2,17	2,54	0,10	-0,94	-0,10	-0,85	0,42	-0,84	-0,04	2
1,06	0,34	2,01	2,22	-0,22	-0,96	0,36	0,27	0,07	0,24	-0,21	2
0,95	0,37	1,60	3,26	-0,25	-2,20	-0,30	0,42	-0,03	0,01	-0,33	2
0,83	0,61	1,29	1,10	-2,23	1,39	0,20	-1,58	-1,29	1,48	-0,41	2
0,98	-0,65	1,43	-0,56	0,76	-0,18	0,88	0,31	0,31	-0,01	0,36	2
0,83	0,30	1,85	1,89	-0,54	-0,98	0,82	1,39	-0,28	1,33	-0,37	2
-0,64	-2,16	-1,74	-1,29	-1,47	-0,31	-0,68	0,63	0,56	0,26	-0,14	3
-0,80	-1,91	-2,59	-1,10	-1,97	-0,90	-0,63	1,33	0,19	0,84	0,23	3
0,88	-0,01	-0,09	-2,60	-0,27	0,09	-1,70	0,41	0,85	0,21	0,21	3
1,90	-2,11	-0,63	-0,76	-0,66	0,37	0,63	0,00	1,39	0,09	0,34	4
2,37	2,34	-0,31	0,06	-0,12	-0,43	-2,34	-0,29	0,02	0,83	-0,29	4
2,33	0,33	0,95	-1,16	0,97	-0,75	-2,24	-0,67	0,28	0,00	-0,28	4
1,99	-1,24	0,59	-0,03	-2,45	0,49	0,56	0,03	1,68	0,08	-1,10	4
2,44	0,29	1,37	-2,21	0,99	0,49	-1,58	-0,82	0,38	0,23	-0,16	4
2,62	-0,12	1,87	0,42	-0,66	-0,09	0,60	0,30	1,60	0,27	1,10	4
1,82	-0,23	0,18	-1,52	-0,91	-0,49	0,38	-0,69	-1,43	0,20	1,59	4
1,82	-0,23	0,18	-1,52	-0,91	-0,49	0,38	-0,69	-1,43	0,20	1,59	4
2,10	-1,51	0,72	-0,76	1,16	-0,26	1,22	0,15	-0,44	-0,61	0,00	4
3,41	1,07	-3,25	2,64	-1,33	-0,03	-0,26	-1,10	-0,34	-0,72	-0,05	4
4,95	3,56	0,29	-0,52	0,53	0,38	1,23	2,65	-0,30	-1,61	0,87	4
3,93	0,99	-3,26	1,44	-0,69	1,23	-0,01	-0,95	0,97	-0,34	0,33	4
1,63	0,07	1,60	-0,88	0,95	0,19	0,45	0,79	0,37	-0,28	-0,42	4
3,89	3,09	-2,73	0,40	1,30	2,42	-0,29	0,91	-0,59	0,17	-1,35	4
1,78	-1,92	0,30	-1,54	-0,78	-0,53	0,51	-0,47	-1,15	-1,16	-0,97	4
1,78	-1,92	0,30	-1,54	-0,78	-0,53	0,51	-0,47	-1,15	-1,16	-0,97	4

TABELA 5.2 - COMPONENTES PRINCIPAIS DAS 51 CASAS

continua												G
1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	9ª	10ª	11ª	12ª	
1,94	-1,59	-0,51	-0,55	1,20	-0,68	0,64	2,52	1,19	0,75	-0,61	-0,10	1
0,01	-0,87	-0,79	1,08	-1,16	0,57	1,08	-0,15	-0,52	1,15	-0,55	0,36	1
1,62	-0,66	-0,48	-1,35	-1,10	-0,31	0,69	0,11	1,19	-0,02	1,39	-0,16	1
2,45	-3,13	-0,56	1,09	0,50	-0,05	-1,90	0,39	0,84	-0,81	-0,07	-0,29	1
0,69	-1,97	-1,22	-0,55	-0,41	-1,74	1,18	0,12	-0,45	-0,08	-0,33	-0,82	1
0,55	-1,97	-0,55	-0,09	2,62	-0,24	-0,55	1,73	-0,17	-0,22	-0,19	0,66	1
0,08	-3,54	-0,14	1,74	1,95	-0,07	-0,54	0,61	-0,32	1,05	-0,07	0,31	1

conclusão												G
1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	9ª	10ª	11ª	12ª	
1,84	-1,64	-0,18	-1,59	-0,30	-0,33	0,71	-0,31	0,54	-1,03	0,97	-0,01	1
0,99	-1,58	-0,67	0,63	-1,15	-0,54	1,04	0,11	-0,81	0,83	-0,05	-0,46	1
1,92	-1,59	-0,31	0,45	-0,96	-1,28	0,21	-0,25	1,07	-0,95	1,06	0,81	1
-0,11	-0,88	0,60	0,28	1,96	0,02	1,52	-0,68	0,07	-0,12	0,89	0,95	1
-0,38	-2,30	0,41	0,20	0,31	-0,05	0,99	-1,05	-1,13	0,01	-0,14	0,21	1
-0,44	-1,59	-1,93	-0,46	-1,69	0,00	-0,55	-0,95	0,31	-0,18	0,05	0,18	1
3,62	-1,84	0,00	-0,45	0,69	0,84	-2,86	-0,57	-0,12	-0,08	-0,79	-1,13	2
3,87	-0,02	0,44	-0,41	-0,36	1,94	-1,59	1,43	0,68	1,25	-0,17	0,62	2
4,17	0,53	0,21	-1,01	-0,40	0,63	0,52	-0,55	0,85	-0,57	0,45	-0,89	2
3,85	1,20	1,41	0,52	0,59	0,52	0,57	-0,76	-0,80	-0,07	-0,54	-0,36	2
2,69	-0,11	0,40	-1,83	-0,39	0,18	-0,39	-0,76	-0,27	-0,82	0,32	0,26	2
2,22	0,82	-0,41	-0,81	0,68	-0,63	0,50	-0,11	-1,09	0,46	0,27	-0,75	2
2,13	0,12	0,54	-0,80	0,50	-0,44	1,63	0,45	-0,32	-0,27	0,04	0,19	2
1,72	1,68	-0,09	-2,60	1,57	-0,39	-0,01	-0,42	-1,45	-0,02	-0,12	-0,68	2
2,87	2,00	0,44	-0,45	-0,53	0,17	0,66	0,49	-0,73	0,66	-0,51	0,32	2
2,73	2,67	0,53	0,41	-0,47	-0,86	-0,11	0,41	-0,41	0,11	-0,42	1,09	2
3,12	0,82	-0,36	-1,25	-1,43	1,32	-0,08	-0,06	-0,04	0,03	-0,02	0,40	2
3,78	2,54	0,90	1,79	-0,18	-0,41	-0,41	-0,56	-0,33	0,03	-0,54	0,27	2
3,06	2,57	0,67	1,67	-0,17	-0,53	-0,56	-0,60	0,04	0,15	-0,51	0,42	2
-1,50	-0,69	0,30	1,05	0,54	0,72	0,12	-0,22	-1,52	-0,78	0,90	-0,24	3
-1,18	-1,05	1,05	0,19	-1,15	2,62	0,14	-0,36	0,68	-1,47	-0,12	0,30	3
-1,58	-1,63	0,06	0,77	-0,09	-0,82	0,24	-1,30	0,35	-0,17	-1,58	0,38	3
-2,85	0,31	1,03	-1,08	-0,16	0,52	0,33	0,41	-0,10	-0,94	-1,01	0,81	3
-3,08	0,05	-0,19	-0,14	1,03	0,97	0,03	-0,23	-0,08	-1,48	-1,02	-0,02	3
-1,49	-1,07	3,65	-0,29	0,05	-1,12	-0,19	-0,99	0,43	1,11	0,56	-1,18	3
-2,97	-0,02	3,75	-0,69	-0,73	0,03	-0,50	0,58	-0,11	0,43	0,35	-0,30	3
-2,24	-0,09	0,42	0,19	-1,37	0,49	0,70	0,78	0,20	0,14	-1,13	0,45	3
-1,84	-1,09	1,40	0,09	0,11	-1,08	0,19	-0,94	0,02	-0,06	-0,70	1,63	3
-2,75	-0,06	3,05	0,97	-0,67	-1,88	-1,42	-0,03	0,51	-0,47	0,15	-1,04	3
-4,38	0,87	2,80	-1,85	-0,33	0,28	-0,78	1,42	-0,83	0,70	0,79	0,64	3
-0,74	-0,66	-0,56	2,25	-1,76	0,08	0,19	0,57	-1,17	-0,47	-0,28	-0,17	4
-2,02	0,40	-2,04	0,50	0,63	0,15	-0,09	-0,60	-0,36	1,14	0,30	-0,28	4
-1,20	0,51	-1,20	1,03	-0,76	0,28	0,15	0,66	-1,04	-0,04	0,69	-0,97	4
-1,47	1,31	-0,11	-0,35	2,00	0,94	1,29	-0,30	1,16	0,85	0,56	0,22	4
-0,78	1,56	0,19	1,94	0,07	1,68	0,95	0,14	1,41	0,08	-0,12	-1,27	4
-0,53	3,05	-1,09	0,82	-0,24	-1,71	-0,80	1,50	0,15	-1,19	1,18	0,12	4
-3,00	1,30	-1,38	-0,41	1,69	1,46	-1,15	-0,56	-1,50	-0,66	0,46	-0,04	4
-2,40	2,36	-1,71	-1,89	-0,48	-1,05	-0,54	-0,44	1,55	0,38	-1,79	-0,22	4
-1,75	0,57	-0,05	1,36	-0,64	1,31	1,05	0,55	0,73	0,61	0,33	-1,02	4
-1,44	2,38	-0,30	1,10	2,50	-0,18	0,04	-1,10	2,03	-0,09	0,46	0,21	4
-1,36	1,85	-1,48	1,21	-0,29	-0,75	-0,17	1,23	-0,31	-1,24	0,43	0,23	4
-1,85	1,25	-2,00	0,29	-0,19	-0,98	-1,29	-1,07	-0,20	1,22	0,25	0,08	4
-4,37	-0,36	-2,00	-2,13	-0,18	-0,24	0,28	0,62	0,03	-0,27	-0,97	-0,96	4
-2,19	-0,74	-1,89	-0,62	-1,44	0,62	-1,19	-0,91	0,16	1,43	1,48	1,21	4

TABELA 5.3 - COMPONENTES PRINCIPAIS DOS 24 TERRENOS

1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	6 ^a	7 ^a	8 ^a	G
-2,80	-1,53	-1,55	-1,07	-0,41	0,94	-0,54	-0,43	1
-0,78	-1,50	0,28	0,78	0,26	0,38	0,61	0,82	1
-1,40	-2,98	-1,76	-0,10	0,01	0,14	0,28	-0,01	1
-0,78	0,05	2,44	-1,79	-0,10	0,75	-0,42	0,63	1
-0,58	-1,54	0,48	0,37	0,16	-0,03	0,51	0,73	1
-1,16	-3,07	-0,95	-1,50	-0,58	-0,94	-0,05	-0,23	1
1,40	0,87	-2,46	1,41	-0,12	-0,43	-0,92	1,04	2
4,32	0,70	-1,03	-1,04	-0,90	0,66	0,84	0,11	2
4,26	-0,22	-0,49	-1,13	0,38	1,19	0,19	-0,61	2
0,17	0,71	0,39	0,27	-1,68	-1,40	0,04	-0,50	2
2,40	-0,39	0,46	-0,01	0,29	-0,29	-1,74	0,41	2
0,23	-0,67	0,15	1,40	0,74	0,16	0,22	-0,21	3
0,26	-1,53	1,60	-0,23	-0,36	-0,53	0,13	0,30	3
1,16	-0,50	1,32	0,22	0,43	-0,86	-0,14	-0,65	3
0,53	-0,87	1,05	0,10	-0,02	-0,86	-0,19	-0,47	3
0,54	-0,88	0,92	1,12	0,98	0,46	0,63	0,42	3
-0,38	0,75	-0,15	1,60	-1,53	-0,12	0,29	-0,29	4
-1,41	1,21	0,05	0,06	0,97	0,72	-0,34	-0,46	4
-0,75	0,82	-0,23	0,64	1,19	0,23	-0,11	-0,62	4
-0,48	1,62	0,31	1,04	-1,11	0,30	0,18	-0,25	4
-0,58	2,97	-1,04	-1,70	1,59	-1,92	0,69	0,37	4
-0,75	0,82	-0,23	0,64	1,19	0,23	-0,11	-0,62	4
-1,98	2,46	-0,18	-0,38	-0,61	0,70	0,04	0,41	4
-1,44	2,70	0,59	-0,73	-0,74	0,51	-0,11	0,13	4

5.3 PREÇOS OBTIDOS NO MODELO PROPOSTO

Conforme já mencionado, após a obtenção dos agrupamentos através da Rede de Kohonen de cada tipo de imóvel avaliado e da eliminação do problema clássico da multicolinearidade dos dados por meio da Análise de Componentes Principais, em cada grupo de imóvel: apartamentos, casas e terrenos, foi aplicada uma Rede Neural Multicamada com a utilização do algoritmo *back-propagation*.

Os melhores resultados obtidos com o modelo proposto estão descritos nas tabelas 5.4 a 5.6. Nestas tabelas *EAX* representa o módulo do erro absoluto do valor do imóvel, que pode ser encontrado pela equação (5.1), sendo x o valor do imóvel e x' o valor calculado pelo modelo.

O critério para determinar o melhor resultado foi encontrar a configuração cuja soma dos módulos do erro absoluto no valor do imóvel fosse a menor. Este também será o critério

adotado para fazer a comparação dos resultados entre a abordagem deste trabalho e dos trabalhos de Alves (2005) e Braulio (2005).

$$EAX = |x - x'| \quad (5.1)$$

Vale lembrar que os critérios de parada fixados são: número máximo de iterações (2000 iterações), erro máximo permitido (10^{-5}) para o conjunto de treinamento ou o número de neurônios na camada escondida (no máximo 50 neurônios).

TABELA 5.4 - MELHOR RESULTADO PARA OS PREÇOS DOS APARTAMENTOS

					continua
PREÇO MERCADO	PREÇO MODELO	ERRO (%)	EAX	GRUPO	
30.000,00	29.999,72	0,00	0,28	1	
40.000,00	40.000,04	0,00	0,04	1	
40.000,00	33.849,67	15,38	6.150,33	1	
50.000,00	49.997,20	0,01	2,80	1	
60.000,00	59.999,94	0,00	0,06	1	
65.000,00	65.008,53	-0,01	8,53	1	
65.000,00	70.263,44	-8,10	5.263,44	1	
65.000,00	64.997,52	0,00	2,48	2	
68.000,00	68.000,03	0,00	0,03	1	
70.000,00	70.000,05	0,00	0,05	1	
71.000,00	71.000,25	0,00	0,25	2	
80.000,00	79.998,63	0,00	1,37	3	
85.000,00	85.004,23	0,00	4,23	3	
90.000,00	89.994,60	0,01	5,40	1	
90.000,00	91.707,20	-1,90	1.707,20	1	
90.000,00	110.288,10	-22,54	20.288,10	2	
93.000,00	93.000,01	0,00	0,01	2	
95.000,00	95.000,05	0,00	0,05	1	
100.000,00	99.999,97	0,00	0,03	1	
100.000,00	100.000,12	0,00	0,12	1	
110.000,00	101.789,62	7,46	8.210,38	1	
110.000,00	58.524,40	46,80	51.475,60	3	
115.000,00	115.006,98	-0,01	6,98	2	
115.000,00	115.000,11	0,00	0,11	4	
120.000,00	119.999,99	0,00	0,01	2	
120.000,00	115.047,91	4,13	4.952,09	2	
120.000,00	120.000,01	0,00	0,01	4	
120.000,00	132.388,11	-10,32	12.388,11	4	
130.000,00	130.000,00	0,00	0,00	2	
140.000,00	139.995,02	0,00	4,98	2	
140.000,00	140.000,10	0,00	0,10	2	
150.000,00	150.000,30	0,00	0,30	4	

				conclusão
PREÇO MERCADO	PREÇO MODELO	ERRO (%)	EAX	GRUPO
150.000,00	149.999,73	0,00	0,27	4
170.000,00	170.000,02	0,00	0,02	4
180.000,00	179.999,92	0,00	0,08	4
180.000,00	179.999,92	0,00	0,08	4
200.000,00	199.999,98	0,00	0,02	4
210.000,00	210.000,02	0,00	0,02	4
220.000,00	259.532,35	-17,97	39.532,35	4
250.000,00	249.992,26	0,00	7,74	4
250.000,00	250.000,01	0,00	0,01	4
250.000,00	249.999,99	0,00	0,01	4
250.000,00	250.007,73	0,00	7,73	4
250.000,00	250.007,73	0,00	7,73	4
			$\sum EAX = 150.029,54$	

Se fosse considerada uma tolerância de 10% para o erro no preço de cada apartamento, haveria apenas cinco (11,36%) dos apartamentos avaliados fora desta tolerância, o que pode ser considerado um resultado muito bom, com um acerto de 88,64%.

TABELA 5.5 - MELHOR RESULTADO PARA OS PREÇOS DAS CASAS

				continua
PREÇO MERCADO	PREÇO MODELO	ERRO (%)	EAX	GRUPO
15.000,00	15.007,40	-0,05	7,40	1
15.000,00	15.004,63	-0,03	4,63	1
28.000,00	28.005,90	-0,02	5,90	2
28.000,00	28.002,68	-0,01	2,68	2
30.000,00	29.996,71	0,01	3,29	1
30.000,00	42.706,36	-42,35	12.706,36	1
30.000,00	29.999,94	0,00	0,06	2
30.000,00	30.191,84	-0,64	191,84	2
35.000,00	34.992,46	0,022	7,54	2
40.000,00	39.999,92	0,00	0,08	2
40.000,00	37.918,09	5,20	2.081,91	2
45.000,00	45.000,01	-0,00	0,01	1
45.000,00	45.000,05	-0,00	0,05	2
45.000,00	45.000,12	-0,00	0,12	2
50.000,00	49.998,69	0,00	1,31	1
50.000,00	50.000,07	-0,00	0,07	1
50.000,00	46.281,25	7,44	3.718,75	1
60.000,00	59.996,72	0,01	3,28	1
60.000,00	60.000,04	-0,00	0,04	2
65.000,00	64.999,91	0,00	0,09	1
70.000,00	70.000,07	-0,00	0,07	1

				conclusão
PREÇO MERCADO	PREÇO MODELO	ERRO (%)	EAX	GRUPO
70.000,00	82.094,09	-17,27	12.094,09	1
85.000,00	84.999,99	0,00	0,01	3
90.000,00	90.000,10	-0,00	0,10	3
90.000,00	179.584,25	-99,54	89.584,25	3
90.000,00	90.000,09	-0,00	0,09	4
95.000,00	34.307,65	63,89	60.692,35	2
98.000,00	98.007,85	-0,01	7,85	4
100.000,00	100.000,06	-0,00	0,06	1
100.000,00	92.135,65	7,86	7.864,35	4
110.000,00	110.000,00	0,00	0,00	4
120.000,00	120.000,24	-0,00	0,24	4
130.000,00	130.000,20	-0,00	0,20	3
130.000,00	130.006,53	-0,01	6,53	3
130.000,00	129.993,96	0,00	6,04	3
135.000,00	129.894,29	3,78	5.105,71	3
140.000,00	140.000,17	-0,00	0,17	2
140.000,00	140.000,20	-0,00	0,20	3
150.000,00	149.999,54	0,00	0,46	2
150.000,00	149.999,95	0,00	0,05	3
150.000,00	149.993,14	0,00	6,86	4
150.000,00	150.000,01	-0,00	0,01	4
150.000,00	149.999,56	0,00	0,44	4
150.000,00	111.320,39	25,79	38.679,61	4
160.000,00	160.000,04	-0,00	0,04	4
160.000,00	142.039,29	11,23	17.960,71	4
165.000,00	164.999,94	0,00	0,06	4
180.000,00	179.999,21	0,00	0,79	3
180.000,00	180.000,00	0,00	0,00	4
185.000,00	129.760,83	29,86	55.239,17	3
290.000,00	290.000,11	-0,00	0,11	4
		$\sum EAX = 305.986,03$		

Considerando-se uma tolerância de 10% para cada casa, resultaria em sete (13,73%) das casas com o preço predito fora dos padrões. O acerto de 86,27% dentro da tolerância pode ser considerado bom, em virtude do tamanho da amostra ser pequeno.

TABELA 5.6 - MELHOR RESULTADO PARA OS PREÇOS DOS TERRENOS

				continua
PREÇO MERCADO	PREÇO MODELO	ERRO (%)	EAX	GRUPO
100.000,00	100.000,77	0,00	0,77	1
47.000,00	47.000,61	0,00	0,61	1
43.000,00	42.996,74	0,01	3,26	1

				conclusão
PREÇO MERCADO	PREÇO MODELO	ERRO (%)	EAX	GRUPO
50.000,00	49.998,13	0,00	1,87	1
145.000,00	144.998,79	0,00	1,21	1
38.000,00	35.939,89	5,42	2.060,11	1
25.000,00	25.000,04	0,00	0,04	2
15.000,00	14.995,54	0,03	4,46	2
15.000,00	15.004,48	-0,03	4,48	2
60.000,00	60.000,01	0,00	0,01	2
13.000,00	14.995,75	-15,35	1.995,75	2
33.000,00	32.999,85	0,00	0,15	3
30.000,00	30.000,33	0,00	0,33	3
38.000,00	37.993,95	0,02	6,05	3
25.000,00	25.001,10	0,00	1,10	3
30.000,00	20.544,74	31,52	9.455,26	3
40.000,00	40.000,48	0,00	0,48	4
100.000,00	99.996,68	0,00	3,32	4
140.000,00	139.998,25	0,00	1,75	4
90.000,00	89.995,27	0,01	4,73	4
70.000,00	69.995,29	0,01	4,71	4
300.000,00	299.999,80	0,00	0,20	4
90.000,00	89.995,27	0,01	4,73	4
250.000,00	276.436,56	-10,57	26.436,56	4
$\sum EAX = 39.991,94$				

Analisando o grupo dos terrenos de acordo com a tolerância de 10%, nota-se que apenas três (12,5%) dos terrenos foram avaliados “inadequadamente” pelo modelo, ou seja, houve um acerto de 87,5%, o que pode ser considerado um resultado aceitável, pois se trata da menor das amostras analisadas, com apenas 24 observações.

Nota-se que em todos os tipos de imóveis (apartamentos, casas e terrenos) alguns elementos “destoam” excessivamente do restante amostra e isto pode indicar duas situações: o imóvel foi avaliado incorretamente ou, então, o imóvel possui alguma(s) característica(s) particular(es) que o diferencia do restante dos imóveis (*outlier*). Como nenhuma técnica de exclusão de *outliers* foi utilizada, não se pode afirmar qual das duas situações ocorreu.

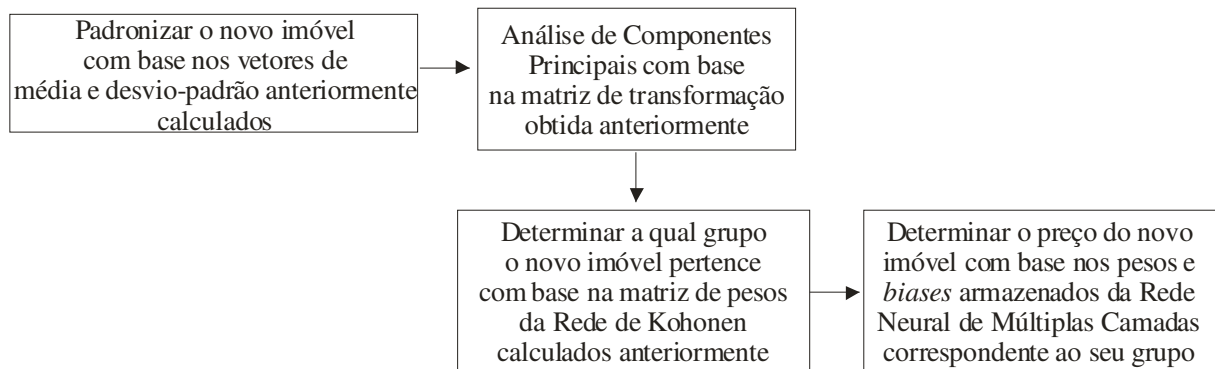
5.4 CÁLCULO DO PREÇO DE NOVOS IMÓVEIS

Conforme descrito no capítulo quatro, o sistema desenvolvido possui um módulo para o cálculo do preço de um novo imóvel tomando como base os dados apresentados durante a execução do programa AIRNA, a figura 5.1 descreve a seqüência para este módulo. Contudo,

não foi possível avaliar satisfatoriamente este módulo, pois não existiam “novos” imóveis para que fosse feita a determinação do preço.

Basicamente, o módulo que calcula o preço de novos imóveis aproveita as informações geradas pela rede durante a execução anterior, aproveitando o vetor de médias e desvio-padrão, os pesos da rede de Kohonen, a matriz de transformação da Análise de Componentes Principais e os pesos e vícios da Rede de Múltiplas Camadas.

FIGURA 5.1 - CÁLCULO DE PREÇOS DE NOVOS IMÓVEIS



5.5 COMPARAÇÕES

Conforme mencionado anteriormente, um dos objetivos deste trabalho é o de comparar o desempenho da abordagem via Redes Neurais com a abordagem por métodos estatísticos multivariados propostos por Alves (2005) e Braulio (2005). Os melhores resultados obtidos no modelo proposto serão comparados com os melhores resultados obtidos nos referidos trabalhos.

As tabelas 5.7 a 5.9 mostram a comparação entre as abordagens no que diz respeito ao preço predito pelo modelo. Na primeira coluna de cada tabela está o preço de mercado do imóvel e nas colunas subsequentes o preços preditos pelo modelo proposto neste trabalho, o preço predito (melhor) em Alves (2005) e, finalmente, o preço predito em Braulio (2005), nesta ordem, com os seus respectivos valores para *EAX*.

O critério utilizado para determinar o melhor resultado entre as abordagens foi o mesmo utilizado anteriormente para determinar a melhor configuração do modelo proposto neste trabalho, ou seja, a menor soma dos módulos dos erros absolutos no valor de cada imóvel.

						conclusão
Mercado	Modelo	EAX	Alves	EAX	Braulio	EAX
250.000,00	250.000,01	0,01	160.269,20	89.730,80	249.772,00	228,00
250.000,00	249.999,99	0,01	250.538,90	538,90	249.481,00	519,00
250.000,00	250.007,73	7,73	227.360,70	22.639,30	249.481,00	519,00
250.000,00	250.007,73	7,73	228.129,50	21.870,50	249.969,00	31,00
		$\sum EAX = 150.029,54$			$\sum EAX = 598.865,53$	$\sum EAX = 171.625,80$

TABELA 5.8 - COMPARAÇÃO REALIZADA PARA CASAS

						continua
Mercado	Modelo	EAX	Alves	EAX	Braulio	EAX
15.000,00	15.007,40	7,40	15.791,10	791,10	18.226,70	3.226,70
15.000,00	15.004,63	4,63	15.620,85	620,85	27.793,40	12.793,40
28.000,00	28.005,90	5,90	24.752,10	3.247,90	28.089,30	89,30
28.000,00	28.002,68	2,68	26.780,53	1.219,47	23.566,90	4.433,10
30.000,00	29.996,71	3,29	29.620,70	379,30	39.095,00	9.095,00
30.000,00	42.706,36	12.706,36	29.383,03	616,97	34.269,50	4.269,50
30.000,00	29.999,94	0,06	12.861,83	17.138,17	17.122,80	12.877,20
30.000,00	30.191,84	191,84	30.220,23	220,23	42.142,90	12.142,90
35.000,00	34.992,46	7,54	38.939,61	3.939,61	42.052,30	7.052,30
40.000,00	39.999,92	0,08	44.249,19	4.249,19	45.914,80	5.914,80
40.000,00	37.918,09	2.081,91	41.536,25	1.536,25	41.513,20	1.513,20
45.000,00	45.000,01	0,01	46.174,35	1.174,35	47.086,40	2.086,40
45.000,00	45.000,05	0,05	42.827,20	2.172,80	43.577,80	1.422,20
45.000,00	45.000,12	0,12	51.735,79	6.735,79	38.064,40	6.935,60
50.000,00	49.998,69	1,31	18.330,54	31.669,46	47.594,80	2.405,20
50.000,00	50.000,07	0,07	50.693,19	693,19	28.207,80	21.792,20
50.000,00	46.281,25	3.718,75	51.443,33	1.443,33	59.734,10	9.734,10
60.000,00	59.996,72	3,28	62.882,68	2.882,68	63.071,00	3.071,00
60.000,00	60.000,04	0,04	64.173,03	4.173,03	87.189,10	27.189,10
65.000,00	64.999,91	0,09	65.899,25	899,25	76.013,70	11.013,70
70.000,00	70.000,07	0,07	83.244,11	13.244,11	78.452,46	8.452,46
70.000,00	82.094,09	12.094,09	67.648,81	2.351,19	101.436,00	31.436,00
85.000,00	84.999,99	0,01	82.956,91	2.043,09	86.798,71	1.798,71
90.000,00	90.000,10	0,10	90.366,90	366,90	89.916,90	83,10
90.000,00	179.584,25	89.584,25	90.626,75	626,75	85.147,90	4.852,10
90.000,00	90.000,09	0,09	87.620,67	2.379,33	94.947,40	4.947,40
95.000,00	34.307,65	60.692,35	94.460,51	539,49	105.746,00	10.746,00
98.000,00	98.007,85	7,85	139.043,20	41.043,20	102.133,00	4.133,00
100.000,00	100.000,06	0,06	100.432,40	432,40	94.574,60	5.425,40

						conclusão
Mercado	Modelo	EAX	Alves	EAX	Braulio	EAX
100.000,00	92.135,65	7.864,35	127.596,00	27.596,00	111.653,00	11.653,00
110.000,00	110.000,00	0,00	104.565,00	5.435,00	120.076,00	10.076,00
120.000,00	120.000,24	0,24	110.594,00	9.406,00	141.580,00	21.580,00
130.000,00	130.000,20	0,20	142.401,60	12.401,60	132.501,00	2.501,00
130.000,00	130.006,53	6,53	117.345,10	12.654,90	119.758,00	10.242,00
130.000,00	129.993,96	6,04	139.631,10	9.631,10	153.964,00	23.964,00
135.000,00	129.894,29	5.105,71	159.128,00	24.128,00	131.932,00	3.068,00
140.000,00	140.000,17	0,17	132.490,30	7.509,70	141.855,00	1.855,00
140.000,00	140.000,20	0,20	149.735,80	9.735,80	137.167,00	2.833,00
150.000,00	149.999,54	0,46	151.772,10	1.772,10	155.189,00	5.189,00
150.000,00	149.999,95	0,05	150.006,90	6,90	145.111,00	4.889,00
150.000,00	149.993,14	6,86	164.577,60	14.577,60	146.178,00	3.822,00
150.000,00	150.000,01	0,01	151.414,40	1.414,40	145.609,00	4.391,00
150.000,00	149.999,56	0,44	134.092,50	15.907,50	152.765,00	2.765,00
150.000,00	111.320,39	38.679,61	146.523,20	3.476,80	135.499,00	14.501,00
160.000,00	160.000,04	0,04	165.112,10	5.112,10	158.597,00	1.403,00
160.000,00	142.039,29	17.960,71	153.527,80	6.472,20	162.315,00	2.315,00
165.000,00	164.999,94	0,06	164.211,70	788,30	177.911,00	12.911,00
180.000,00	179.999,21	0,79	181.586,70	1.586,70	177.199,00	2.801,00
180.000,00	180.000,00	0,00	178.216,50	1.783,50	193.211,00	13.211,00
185.000,00	129.760,83	55.239,17	166.023,60	18.976,40	163.823,00	21.177,00
290.000,00	290.000,11	0,11	243.132,70	46.867,30	275.629,00	14.371,00
		$\sum EAX = 305.986,03$	$\sum EAX = 386.069,28$		$\sum EAX = 422.449,07$	

TABELA 5.9 - COMPARAÇÃO REALIZADA PARA TERRENOS

						continua
Mercado	Modelo	EAX	Alves	EAX	Braulio	EAX
13.000,00	14.995,75	1.995,75	8.133,56	4.866,44	14.913,50	1.913,50
15.000,00	14.995,54	4,46	19.406,38	4.406,38	17.519,80	2.519,80
15.000,00	15.004,48	4,48	15.444,18	444,18	14.013,10	986,90
25.000,00	25.000,04	0,04	24.905,15	94,85	21.410,60	3.589,40
25.000,00	25.001,10	1,10	31.039,01	6.039,01	25.314,30	314,30
30.000,00	30.000,33	0,33	34.402,67	4.402,67	33.631,40	3.631,40
30.000,00	20.544,74	9.455,26	27.879,77	2.120,23	25.537,10	4.462,90
33.000,00	32.999,85	0,15	28.632,82	4.367,18	27.897,60	5.102,40
38.000,00	35.939,89	2.060,11	34.325,81	3.674,19	32.633,90	5.366,10
38.000,00	37.993,95	6,05	41.672,51	3.672,51	43.506,30	5.506,30
40.000,00	40.000,48	0,48	37.169,79	2.830,21	40.768,60	768,60
43.000,00	42.996,74	3,26	48.369,13	5.369,13	48.922,40	5.922,40
47.000,00	47.000,61	0,61	42.870,81	4.129,19	50.870,10	3.870,10

						conclusão
Mercado	Modelo	EAX	Alves	EAX	Braulio	EAX
50.000,00	49.998,13	1,87	51.225,11	1.225,11	48.622,40	1.377,60
60.000,00	60.000,01	0,01	51.262,84	8.737,16	51.421,60	8.578,40
70.000,00	69.995,29	4,71	77.048,69	7.048,69	77.704,20	7.704,20
90.000,00	89.995,27	4,73	90.000,00	0,00	88.880,60	1.119,40
90.000,00	89.995,27	4,73	90.000,00	0,00	88.880,60	1.119,40
100.000,00	100.000,77	0,77	103.178,10	3.178,10	104.910,00	4.910,00
100.000,00	99.996,68	3,32	100.000,00	0,00	100.164,00	164,00
140.000,00	139.998,25	1,75	140.000,00	0,00	145.725,00	5.725,00
145.000,00	144.998,79	1,21	140.033,70	4.966,30	137.404,00	7.596,00
250.000,00	276.436,56	26.436,56	250.000,00	0,00	242.536,00	7.464,00
300.000,00	299.999,80	0,20	300.000,00	0,00	303.813,00	3.813,00
	$\sum EAX$	39.991,94	$\sum EAX$	71.571,53	$\sum EAX$	93.525,10

A tabela 5.7 mostrou que a abordagem via Redes Neurais para apartamentos fornece uma soma dos módulos dos erros absolutos de R\$ 150.029,54, enquanto a abordagem de Alves (2005) fornece um erro 299,17% maior (R\$ 598.865,53), uma diferença consideravelmente grande dado o número de apartamentos. Já o resultado obtido por Braulio (2005) foi bem próximo ao obtido no modelo proposto, com uma inferioridade na predição dos preços de apenas 14,39% (R\$ 171.625,80). Contudo, foi possível notar uma superioridade na capacidade preditiva dos preços da abordagem via Redes Neurais.

A tabela 5.8 mostrou a mesma análise de sua antecessora, porém o tipo de imóvel em questão eram as casas residenciais avaliadas. A abordagem via Redes Neurais mostrou uma soma dos módulos dos erros absolutos de R\$ 305.986,03 e, neste caso, a diferença entre as abordagens foi muito menor do que no caso dos apartamentos, pois na abordagem de Alves (2005) a soma em questão foi apenas 26,17% maior (R\$ 386.069,28) e para Braulio (2005), o resultado foi um pouco pior, 38,06% maior (R\$ 422.449,07).

A última das três tabelas que mostram a comparação entre as abordagens, a tabela 5.9, mostrou a análise para os terrenos avaliados. O modelo proposto teve uma somatória dos módulos dos erros absolutos igual a R\$ 39.991,94, resultado que foi consideravelmente superior ao de Alves (2005) que teve um valor 78,96% maior (R\$ 71.571,53) e melhor ainda em relação ao de Braulio (2005), que foi 133,86% maior (R\$ 93.525,10).

Assim sendo, foi possível notar um desempenho consideravelmente superior da abordagem via Redes Neurais em relação às outras abordagens aqui comparadas, para todos os três tipos de imóveis analisados: apartamentos, casas e terrenos.

6 CONSIDERAÇÕES FINAIS

6.1 CONCLUSÕES

As conclusões pertinentes ao trabalho são:

- Levando-se em consideração que as amostras para este trabalho são extremamente limitadas, pois seria “normal” haver, no mínimo, 100 padrões em cada tipo de imóvel, pode-se considerar que nos três casos houve um desempenho aceitável da rede. A rede foi capaz de generalizar razoavelmente, em virtude do tamanho da amostra, os preços dos imóveis: apartamentos, casas e terrenos. Provavelmente, se a amostra fosse maior o desempenho da rede teria sido melhor ainda;
- O tratamento estatístico dos dados constituído pela padronização e Análise de Componentes Principais dos imóveis foi de extrema importância, pois garantiu a não multicolinearidade dos dados e, também, que somente as variáveis significativas fossem apresentadas à rede, garantindo certamente um melhor desempenho da mesma, alguns testes sem utilizar a Análise de Componentes Principais foram efetuados, entretanto, na maioria dos casos a Rede Neural de Múltiplas Camadas sequer convergia;
- Com relação ao sistema AIRNA desenvolvido, também não é recomendável não realizar a Análise de Componentes Principais mesmo que esta ofereça a opção de não fazê-lo, pois assim não há garantias de que o problema da multicolinearidade não afete os resultados;
- Com relação à arquitetura da Rede de Kohonen, não se mostrou recomendável a utilização de uma topologia que não fosse de 2x2 neurônios, pois para topologias diferentes desta, apesar do desempenho no conjunto de treinamento ser muito bom, o desempenho do conjunto de testes ficava altamente prejudicado, pois os conjuntos de treinamento e testes ficavam ainda menores; a rede perde consideravelmente a capacidade de generalização, que já não foi excelente, fato certamente indesejável em qualquer análise de dados com Redes Neurais;
- A metodologia aplicada neste trabalho mostrou um desempenho levemente superior aos trabalhos de Braulio (2005) e Alves (2005), considerando todos os tipos de

imóveis (apartamentos, casas e terrenos), no que diz respeito ao critério adotado, ou seja, a soma dos módulos dos erros absolutos (quanto menor, melhor).

6.2 SUGESTÕES PARA TRABALHOS FUTUROS

1. Aplicar as metodologias propostas em Alves (2005) e Braulio (2005) e a metodologia proposta neste trabalho em amostras maiores de imóveis, comparando os seus resultados, afim de verificar qual proposta apresenta o melhor desempenho;
2. Criar uma interface gráfica, para o sistema desenvolvido, na qual haja mais praticidade para a determinação dos preços de novos imóveis, dada a dificuldade de se efetuar esta operação com a disposição atual;
3. Aplicar outras técnicas de predição como, por exemplo, Árvores de Decisão, Algoritmos Genéticos, entre outros para comparar resultados.

REFERÊNCIAS

ABNT (Associação Brasileira de Normas Técnicas). **Avaliação de imóveis urbanos** (NBR 5676 e NBR 502). Rio de Janeiro: ABNT, 2004.

ALVES, V. **Avaliação de imóveis urbanos baseada em métodos estatísticos multivariados**. Dissertação de Mestrado. UFPR: Curitiba, 2005.

BAPTISTELLA, M. **O uso de redes neurais e regressão múltipla na engenharia de avaliações: determinação dos valores venais de imóveis**. Dissertação de Mestrado. UFPR: Curitiba, 2005.

BRAÚLIO, S. N. **Proposta de uma metodologia para a avaliação de imóveis urbanos baseado em métodos estatísticos multivariados**. Dissertação de Mestrado. UFPR: Curitiba, 2005.

DANTAS, R. A. **Engenharia de avaliações: introdução à metodologia científica**. São Paulo: Pini, 1998.

FAUSETT, L. **Fundamentals of neural networks: architectures, algorithms, and applications**. Upper Saddle River, NJ: Prentice Hall, 1994. 459 p., ill.

FIKER, J. **Avaliação de imóveis urbanos**. 5. ed. São Paulo: Pini, 1997.

FRANCISCO, C. A. C. **Rede de Kohonen: uma ferramenta no estudo das relações tróficas entre espécies de peixes**. Dissertação de Mestrado. UFPR: Curitiba, 2004.

HAYKIN, S. **Redes neurais: princípios e prática**. 2.ed. Tradução de, Paulo Martins Engel. PortoAlegre: Bookman, 2001. 900 p., il.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4. ed. Nova Jersey: Prentice Hall, Inc., 2002.

MOSCOVITCH, S. K. **Qualidade de vida urbana e valores de imóveis: um estudo de caso para Belo Horizonte**. Nova Economia, número especial: 247-279, 1997.

STEINER, Maria Teresinha Arns. **Uma metodologia para o reconhecimento de padrões multivariados com resposta dicotômica**. Tese de Doutorado, Universidade Federal de Santa Catarina – UFSC. Programa de Pós-Graduação em Engenharia de Produção – PPGEP, 1995.

ANEXOS

ANEXO I – MATRIZ DE DADOS REFERENTE AOS APARTAMENTOS

continua

condomínio																						
apto	posição do apto	elevador	garagem	local.	área	pavimento	andar	peças	salas	dormitório	suíte	banheiro	dep. de emp.	dist. escola	dist. hospital	dist. de merc.	acab.	revest. prédio	conservação	idade real	idade aparente	valor (R\$)
1	3	2	1	1	222	15	3	9	1	2	1	1	1	2	2	3	3	4	2	2	2130000	
2	3	1	1	1	162,44	7	1	9	1	2	1	1	1	3	1	3	3	1	5	4	6 85000	
3	3	1	1	1	176	8	2	9	1	2	1	1	1	3	2	3	3	1	5	4	6 80000	
4	3	2	2	1	179,2	14	3	10	2	2	1	1	1	3	2	3	3	4	5	4	6115000	
5	3	2	2	1	279,4	20	3	8	1	2	1	1	1	3	2	3	3	4	5	2	4150000	
6	3	1	1	1	120	12	3	12	2	3	1	2	1	3	2	3	3	2	5	3	4110000	
7	2	1	1	1	220	16	2	12	3	3	1	2	1	2	2	2	3	4	4	2	2120000	
8	3	1	1	1	107	8	1	9	3	2	1	1	0	2	2	2	2	1	3	3	5 68000	
9	2	0	1	1	50	4	2	4	1	1	0	1	0	3	2	3	2	1	4	4	4 40000	
10	3	2	3	1	240	15	3	12	3	2	1	1	1	3	3	3	3	4	4	2	3170000	
11	3	0	1	0	100	3	2	7	2	3	0	2	0	3	1	1	2	1	3	2	3 50000	
12	3	0	1	1	147	6	1	7	1	2	1	1	0	3	2	3	2	2	5	4	5 60000	
13	3	1	1	1	108	7	3	11	2	2	1	1	1	3	3	3	2	4	3	3	3 93000	
14	2	2	2	1	311	16	3	10	3	2	1	1	1	1	1	1	3	4	5	4	4250000	
15	2	1	1	1	132	7	2	8	2	2	1	1	0	3	2	1	2	1	4	5	5 65000	
16	3	2	1	1	107	8	3	8	2	2	1	1	0	3	3	3	2	4	4	3	3 65000	
17	3	2	1	1	107	8	4	8	2	2	1	1	0	3	3	3	2	4	4	3	3 71000	
18	3	2	4	1	330	15	3	17	3	3	1	3	1	2	1	2	2	4	4	3	2220000	
19	3	2	2	1	220	15	3	13	2	3	1	1	1	3	3	3	2	4	4	5	5200000	
20	3	1	1	1	130	6	3	11	2	3	1	1	1	3	3	3	2	3,5	4	3	4 90000	
21	3	2	2	1	164	10	3	11	2	3	1	1	1	3	3	3	2	3,5	4	3	4140000	
22	3	2	2	1	160	15	3	11	2	3	1	2	1	3	3	3	2	4	4	3	4250000	
23	3	2	1	1	374	15	3	14	3	2	1	3	1	1	1	1	2	4	4	4	4250000	
24	3	1	1	1	220	16	4	14	3	3	1	2	1	3	3	3	3	4	4	3	4150000	
25	2	1	1	1	220	16	3	14	3	3	1	2	1	3	3	3	3	4	4	3	4120000	
26	3	1	2	1	180	13	3	13	2	3	1	1	1	3	1	3	3	4	3	4	4180000	
27	3	1	2	1	320	13	3	11	1	3	1	1	1	3	3	3	3	4	4	5	5250000	

conclusão																						
apto	posição do apto	elevador	garagem	local.	área	pavimento	andar	peças	salas	dormitório	suíte	banheiro	dep. de emp.	dist. escola	dist. hospital	dist. de merc.	acab.	revest. prédio	conservação	idade real	idade aparente	valor (R\$)
28	3	1	2	1	180		13	3	13	2	3	1	1	1	3	1	3	3	4	3	4	4180000
29	3	1	2	1	320		13	3	11	1	3	1	1	1	3	3	3	3	4	4	5	5250000
30	1	2	2	1	260		14	3	11	2	2	1	1	1	3	3	3	2	4	3	2	2115000
31	2	2	2	1	260		14	2	11	2	2	1	1	1	3	3	3	2	4	3	2	2120000
32	3	2	2	1	260		14	1	11	2	2	1	1	1	3	3	3	2	4	3	2	2140000
33	3	1	1	1	140		8	2	6	1	2	1	1	0	2	2	3	2	4	3	5	5 90000
34	1	1	1	1	130		7	3	7	2	2	1	1	0	3	3	3	2	1	3	4	4 65000
35	1	2	2	1	260		14	1	11	2	2	1	1	1	3	3	3	2	4	3	2	2120000
36	1	2	2	1	310		16	2	11	2	2	1	1	1	1	1	1	3	4	4	4	4210000
37	3	0	1	1	100		4	3	7	1	3	0	2	0	3	3	3	2	1	3	1	1100000
38	3	0	1	1	70		3	2	5	1	2	0	1	0	3	3	3	2	1	3	1	1 70000
39	3	0	1	1	90		3	2	6	1	2	0	2	0	3	3	3	2	1	2	1	1 95000
40	3	1	1	1	38		8	2	4	1	1	0	1	0	3	3	3	2	1	3	3	4 30000
41	3	1	1	1	40		8	3	5	1	2	0	1	0	3	3	3	2	1	3	3	4 40000
42	3	1	1	1	170		7	2	9	2	2	1	1	0	3	3	3	2	4	4	4	6100000
43	1	1	1	1	170		7	1	9	2	2	1	1	0	3	3	3	2	4	4	4	6 90000
44	3	1	1	1	170		7	3	9	2	2	1	1	0	3	3	3	2	4	4	4	6110000

ANEXO II – MATRIZ DE DADOS REFERENTE AS CASAS RESIDENCIAIS

																			continua
casa	bairro	garagem	suíte	banheiro	edícula	dist. mercado	área const.	área terreno	acab.	cobertura	estrutura	conserv.	piscina	dormitório	dep. emp.	lav.	peças	idade aparente	valor (R\$)
1	5	1	1	2	1	3	183	500	1	4	3	1	0	2	1	1	8	2120000	
2	4	1	1	3	0	2	216	977	1	4	3	1	0	3	1	1	13	2160000	
3	4	1	1	3	0	2	155	420	2	4	3	1	0	3	1	1	12	2110000	
4	2	1	1	2	0	1	170	490	2	4	3	1	0	2	0	1	7	3 70000	
5	4	1	1	2	1	1	160	480	2	4	3	3	0	2	0	1	9	3 85000	
6	5	1	1	2	1	3	160	500	2	4	3	3	0	3	0	1	8	3100000	
7	5	1	1	2	1	3	134	1000	2	4	3	2	0	4	0	1	10	4160000	
8	3	1	0	1	0	3	70	300	1	2	1	1	0	3	0	0	5	1 30000	
9	3	0	0	1	0	2	198	350	1	1	1	3	0	2	0	0	5	3 28000	
10	3	1	0	1	0	2	113	400	2	2	1	2	0	3	0	0	8	2 35000	
11	5	1	0	2	1	3	238	1200	2	4	3	2	0	4	0	1	9	1150000	
12	3	1	0	2	0	2	158	315	2	1	1	2	0	4	0	1	12	1 45000	
13	3	1	0	1	0	3	124	300	2	4	3	2	0	3	0	0	8	1 30000	
14	5	0	0	1	0	3	95	340	2	2	1	1	0	2	0	0	6	1 28000	
15	4	1	1	1	1	3	187	490	3	4	3	2	0	2	1	1	12	4140000	
16	4	1	1	1	1	1	242	490	3	4	3	2	0	3	1	1	13	3130000	
17	3	0	0	2	0	1	100	480	2	4	3	2	0	3	0	1	8	5 50000	
18	3	1	1	1	0	1	180	786	3	4	3	3	0	2	1	1	10	3150000	
19	5	1	1	3	1	1	380	480	2	3	3	2	0	3	0	1	14	5150000	
20	4	1	1	1	1	1	240	490	3	4	3	2	1	2	1	1	11	2135000	
21	3	1	1	1	1	2	184	1000	2	4	3	3	1	2	1	1	13	3180000	
22	3	1	0	2	0	3	140	446	2	1	2	2	0	3	0	1	6	2 40000	
23	5	1	1	1	0	3	400	600	2	4	3	2	0	5	1	1	14	3150000	
24	4	1	0	1	0	3	110	270	2	4	3	3	0	2	0	1	5	4 60000	
25	2	0	0	1	0	3	120	300	2	4	3	1	0	3	0	1	6	3 15000	
26	2	1	0	1	0	3	150	300	2	4	3	3	0	3	0	1	6	5 45000	
27	5	1	1	2	0	3	400	750	2	4	3	3	0	3	0	1	11	3165000	
28	5	1	0	1	0	3	130	300	2	2	1	1	0	3	0	0	5	3 95000	

																			conclusão
casa	bairro	garagem	suíte	banheiro	edícula	dist. mercado	área const.	área terreno	acab.	cobertura	estrutura	conserv.	piscina	dormitório	dep. emp.	lav.	peças	idade aparente	valor (R\$)
29	4	1	1	1	0	3	200	400	2	4	3	3	0	3	0	0	9		6100000
30	4	1	1	2	1	1	244	500	2	4	3	2	0	3	1	1	13		6130000
31	5	1	1	1	1	1	113	300	2	4	3	2	0	2	1	0	9		4 90000
32	5	1	1	2	1	1	220	500	4	4	3	2	1	3	1	1	15		3185000
33	5	1	1	1	1	3	92	650	2	4	3	3	0	2	0	1	7		5 90000
34	5	1	1	1	0	3	123	300	2	4	3	2	0	2	0	1	7		5 15000
35	5	1	0	1	0	3	150	1000	1	2	1	1	0	2	0	1	7		1150000
36	5	1	1	2	1	3	200	400	2	4	4	2	0	2	1	1	7		2150000
37	5	1	0	1	0	3	100	1000	1	1	1	1	0	2	0	1	5		1140000
38	4	1	0	1	0	2	70	490	1	1	1	1	0	2	0	1	5		1 30000
39	5	1	0	1	0	3	100	950	2	2	1	1	0	3	0	1	7		1 60000
40	5	1	0	1	0	3	80	475	2	2	1	1	0	3	0	1	7		1 45000
41	3	0	0	1	0	2	70	600	1	3	3	3	0	2	0	0	5		5 30000
42	3	1	1	1	0	2	180	640	2	4	3	3	0	2	1	1	10		6 90000
43	2	0	1	2	0	2	70	480	2	4	3	3	0	1	0	1	8		5 50000
44	3	1	1	1	0	1	115	250	2	4	3	3	0	2	0	1	8		4 70000
45	3	1	1	1	0	1	160	450	2	4	3	3	1	2	1	1	9		1130000
46	3	1	1	2	1	3	325	225	3	4	4	3	0	4	1	1	16		6180000
47	5	1	1	2	0	3	320	450	3	4	4	3	0	2	0	0	13		4290000
48	2	1	0	1	0	2	116	300	2	3	3	2	0	3	0	0	6		3 50000
49	5	1	1	3	0	3	180	500	2	4	3	3	0	3	0,5	1	12		5 98000
50	2	1	0	1	0	3	100	800	2	3,5	3	2	0	2	0	0	5		4 65000
51	3	1	0	1	0	2	80	390	2	2	3	1	0	3	0	1	6		2 40000

ANEXO III – MATRIZ DE DADOS REFERENTE AOS TERRENOS

terreno	localização	setor comercial	pólo	frente	área do terreno	proteção	plano	inclinado	posição	pavimentação	valor (R\$)
1	6		2	1	2	650	3	0	1	1	100000
2	2		0	0	3	640	3	1	0	2	25000
3	5		0	1	1	500	3	1	0	1	43000
4	3		0	1	2	390	3	2	0	1	33000
5	5		0	1	3	262	3	2	0	2	40000
6	5		2	1	3	1000	3	2	0	1	140000
7	4		1	1	3	950	3	2	0	1	90000
8	4		0	1	1	475	3	0	1	1	38000
9	5		1	1	3	470	3	3	0	2	70000
10	6		3	1	1	500	0	3	0	1	145000
11	2		0	0	2	420	0	3	0	2	15000
12	2		0	0	2	420	0	3	0	1	13000
13	3		2	1	3	2000	0	2	0	2	100000
14	4		1	1	3	950	3	2	0	1	90000
15	6		3	1	3	1000	3	2	0	2	250000
16	5		0	1	1	242	0	2	0	1	38000
17	6		3	1	3	940	2	3	0	2	300000
18	5		0	1	1	500	2	1	0	1	47000
19	5		0	1	3	350	0	2	0	2	60000
20	5		0	1	1	500	0	0	1	1	50000
21	3		0	0	2	336	0	3	0	1	15000
22	3		0	1	2	336	0	3	0	1	30000
23	4		0	1	2	300	0	2	0	1	25000
24	3		0	1	1	450	3	3	0	1	30000